MAPS OF BOUNDED RATIONALITY: A PERSPECTIVE ON INTUITIVE JUDGMENT AND CHOICE

Prize Lecture, December 8, 2002

by

DANIEL KAHNEMAN*

Princeton University, Department of Psychology, Princeton, NJ 08544, USA.

The work cited by the Nobel committee was done jointly with the late Amos Tversky (1937–1996) during a long and unusually close collaboration. Together, we explored the psychology of intuitive beliefs and choices and examined their bounded rationality. This essay presents a current perspective on the three major topics of our joint work: heuristics of judgment, risky choice, and framing effects. In all three domains we studied intuitions – thoughts and preferences that come to mind quickly and without much reflection. I review the older research and some recent developments in light of two ideas that have become central to social-cognitive psychology in the intervening decades: the notion that thoughts differ in a dimension of accessibility – some come to mind much more easily than others – and the distinction between intuitive and deliberate thought processes.

Section 1 distinguishes two generic modes of cognitive function: an intuitive mode in which judgments and decisions are made automatically and rapidly, and a controlled mode, which is deliberate and slower. Section 2 describes the factors that determine the relative accessibility of different judgments and responses. Section 3 explains framing effects in terms of differential salience and accessibility. Section 4 relates prospect theory to the general

* This essay revisits problems that Amos Tversky and I studied together many years ago, and continued to discuss in a conversation that spanned several decades. The article is based on the Nobel lecture, which my daughter Lenore Shoham helped put together. It builds on an analysis of judgment heuristics that was developed in collaboration with Shane Frederick (Kahneman and Frederick, 2002). Shane Frederick, David Krantz, and Daniel Reisberg went well beyond the call of friendly duty in helping with this effort. Craig Fox, Peter McGraw, Daniel Read, David Schkade and Richard Thaler offered many insightful comments and suggestions. Kurt Schoppe provided valuable assistance, and Geoffrey Goodwin and Amir Goren helped with scholarly fact-checking. My research is supported by NSF 285-6086 and by the Woodrow Wilson School for Public and International Affairs at Princeton University. A different version of this article is to appear in the American Economic Review (December 2003).

proposition that changes and differences are more accessible than absolute values. Section 5 reviews an attribute substitution model of heuristic judgment. Section 6 describes a particular family of heuristics, called prototype heuristics. Section 7 concludes with a review of the argument.

1. INTUITION AND ACCESSIBILITY

From its earliest days, the research that Tversky and I conducted was guided by the idea that intuitive judgments occupy a position – perhaps corresponding to evolutionary history – between the automatic operations of perception and the deliberate operations of reasoning. Our first joint article examined systematic errors in the casual statistical judgments of statistically sophisticated researchers (Tversky & Kahneman, 1971). Remarkably, the intuitive judgments of these experts did not conform to statistical principles with which they were thoroughly familiar. In particular, their intuitive statistical inferences and their estimates of statistical power showed a striking lack of sensitivity to the effects of sample size. We were impressed by the persistence of discrepancies between statistical intuition and statistical knowledge, which we observed both in ourselves and in our colleagues. We were also impressed by the fact that significant research decisions, such as the choice of sample size for an experiment, are routinely guided by the flawed intuitions of people who know better. In the terminology that became accepted much later, we held a two-system view, which distinguished intuition from reasoning. Our research focused on errors of intuition, which we studied both for their intrinsic interest and for their value as diagnostic indicators of cognitive mechanisms.

The two-system view

The distinction between intuition and reasoning has been a topic of considerable interest in the intervening decades (among many others, see Epstein, 1994; Hammond, 1996; Jacoby, 1981, 1996; and numerous models collected by Chaiken & Trope, 1999; for comprehensive reviews of intuition, see Hogarth, 2002; Myers, 2002). In particular, the differences between the two modes of thought have been invoked in attempts to organize seemingly contradictory results in studies of judgment under uncertainty (Kahneman & Frederick, 2002; Sloman, 1996, 2002; Stanovich, 1999; Stanovich & West, 2002). There is considerable agreement on the characteristics that distinguish the two types of cognitive processes, which Stanovich and West (2000) labeled System 1 and System 2. The scheme shown in Figure 1 summarizes these characteristics: The operations of System 1 are fast, automatic, effortless, associative, and difficult to control or modify. The operations of System 2 are slower, serial, effortful, and deliberately controlled; they are also relatively flexible and potentially rule-governed. As indicated in Figure 1, the operating characteristics of System 1 are similar to the features of perceptual processes. On the other hand, as Figure 1 also shows, the operations of System 1, like those of System 2, are not restricted to the processing of current stimulation. Intuitive judgments deal with concepts as well as with percepts, and can be evoked by language.

In the model that will be presented here, the perceptual system and the intuitive operations of System 1 generate *impressions* of the attributes of objects of perception and thought. These impressions are not voluntary and need not be verbally explicit. In contrast, *judgments* are always explicit and intentional, whether or not they are overtly expressed. Thus, System 2 is involved in all judgments, whether they originate in impressions or in deliberate reasoning. The label 'intuitive' is applied to judgments that directly reflect impressions. As in several other dual-process models, one of the functions of System 2 is to monitor the quality of both mental operations and overt behavior (Gilbert, 2002; Stanovich & West, 2002). In the anthropomorphic terms that will be used here, the explicit judgments that people make (whether overt or not) are endorsed, at least passively, by System 2. Kahneman and Frederick (2002) suggested that the monitoring is normally quite lax, and allows many intuitive judgments to be expressed, including some that are erroneous.

Shane Frederick (personal communication, April 2003) has used simple puzzles to study cognitive self-monitoring, as in the following example: "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?" Almost everyone reports an initial tendency to answer "10 cents" because the sum $1.10 separates naturally into $1 and 10 cents, and 10 cents is about the right magnitude. Frederick found that many intelligent people yield to this immediate impulse: 50% (47/93) of Princeton students, and 56% (164/293) of students at the University of Michigan gave the wrong answer. Clearly, these respondents offered a response without checking it. The surprisingly high rate of errors in this easy problem illustrates how lightly the output of System 1 is monitored by System 2: people are not accus-

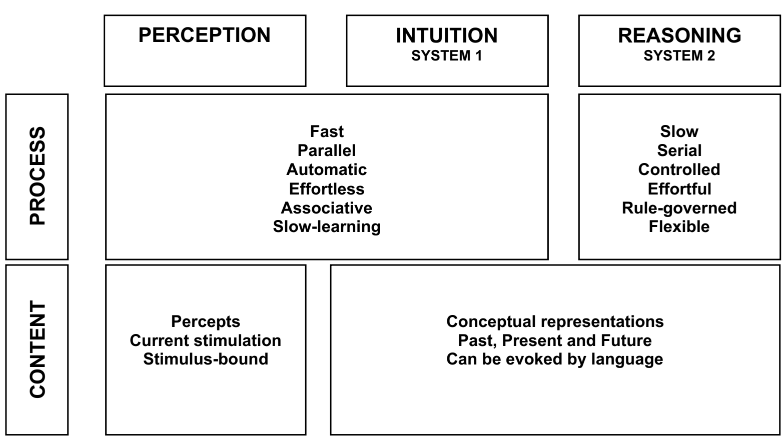


Figure 1.

tomed to thinking hard, and are often content to trust a plausible judgment that quickly comes to mind. Remarkably, errors in this puzzle and in others of the same type were significant predictors of relative indifference to delayed rewards (high discount rates), and of cheating.

The accessibility dimension

The core concept of the present analysis of intuitive judgments and preferences is *accessibility* – the ease with which particular mental contents come to mind (Higgins, 1996). A defining property of intuitive thoughts is that they come to mind spontaneously, like percepts. To understand intuition, then, we must understand why some thoughts are accessible and others are not. The concept of accessibility is applied more broadly in this treatment than in common usage. Category labels, descriptive dimensions (attributes, traits), values of dimensions, all can be described as more or less accessible, for a given individual exposed to a given situation at a particular moment.

For an illustration of differential accessibility, consider Figures 2a and 2b. As we look at the object in Figure 2a, we have immediate impressions of the height of the tower, the area of the top block, and perhaps the volume of the tower. Translating these impressions into units of height or volume requires a deliberate operation, but the impressions themselves are highly accessible. For other attributes, no perceptual impression exists. For example, the total area that the blocks would cover if the tower were dismantled is not perceptually accessible, though it can be estimated by a deliberate procedure, such as multiplying the area of a block by the number of blocks. Of course, the situation is reversed with Figure 2b. Now the blocks are laid out and an impression of total area is immediately accessible, but the height of the tower that could be constructed with these blocks is not.

Some relational properties are accessible. Thus, it is obvious at a glance

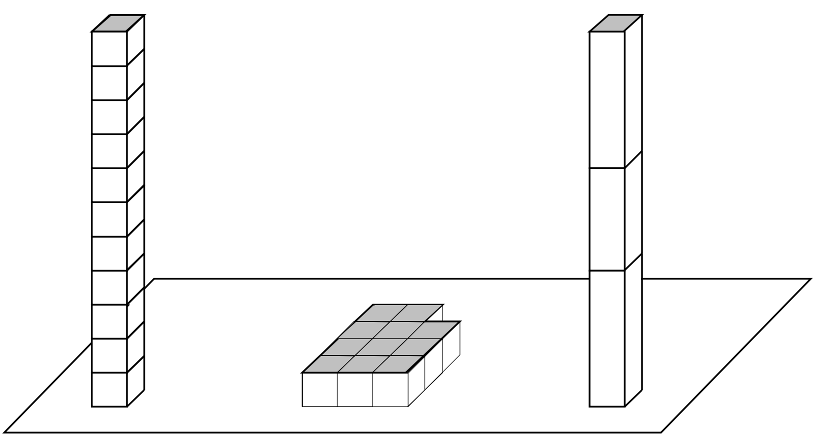


Figure 2a. *Figure 2b.* *Figure 2c.*

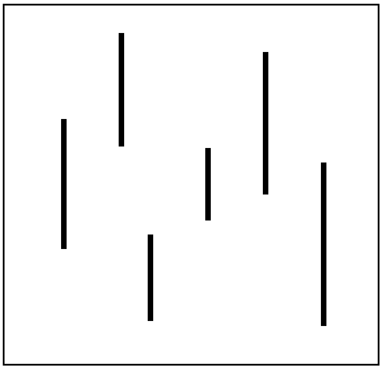


Figure 3.

that Figures 2a and 2c are different, but also that they are more similar to each other than either is to Figure 2b. And some statistical properties of ensembles are accessible, while others are not. For an example, consider the question "What is the average length of the lines in Figure 3?" This question is easy. When a set of objects of the same general kind is presented to an observer – whether simultaneously or successively – a representation of the set is computed automatically, which includes quite precise information about the average (Ariely, 2001; Chong & Treisman, in press). The representation of the prototype is highly accessible, and it has the character of a percept: we form an impression of the typical line without choosing to do so. The only role for System 2 in this task is to map this impression of typical length onto the appropriate scale. In contrast, the answer to the question "What is the total length of the lines in the display?" does not come to mind without considerable effort.

These perceptual examples serve to establish a dimension of accessibility. At one end of this dimension we find operations that have the characteristics of perception and of the intuitive System 1: they are rapid, automatic, and effortless. At the other end are slow, serial and effortful operations that people need a special reason to undertake. Accessibility is a continuum, not a dichotomy, and some effortful operations demand more effort than others. The acquisition of skill selectively increases the accessibility of useful responses and of productive ways to organize information. The master chess player does not see the same board as the novice, and the skill of visualizing the tower that could be built from an array of blocks could surely be improved by prolonged practice.

Determinants of accessibility

As it is used here, the concept of accessibility subsumes the notions of stimulus salience, selective attention, and response activation or priming. The different aspects and elements of a situation, the different objects in a scene, and the different attributes of an object – all can be more or less accessible. What becomes accessible in any particular situation is mainly determined, of course, by the actual properties of the object of judgment: it is easier to see a tower in Figure 2a than in Figure 2b, because the tower in the latter is only virtual. Physical salience also determines accessibility: if a large green letter and a small blue letter are shown at the same time, 'green' will come to mind first. However, salience can be overcome by deliberate attention: an instruction to look for the smaller letter will enhance the accessibility of all its features. Motivationally relevant and emotionally arousing stimuli spontaneously attract attention. All the features of an arousing stimulus become accessible, including those that are not linked to its motivational or emotional significance. This fact is known, of course, to the designers of billboards.

The perceptual effects of salience and of spontaneous and voluntary attention have counterparts in the processing of more abstract stimuli. For example, the statements 'Team A beat team B' and 'Team B lost to team A' convey the same information. Because each sentence draws attention to its subject, however, the two versions make different thoughts accessible. Accessibility also reflects temporary states of priming and associative activation, as well as enduring operating characteristics of the perceptual and cognitive systems. For example, the mention of a familiar social category temporarily increases the accessibility of the traits associated with the category stereotype, as indicated by a lowered threshold for recognizing manifestations of these traits (Higgins, 1996; for a review, see Fiske, 1998). And the "hot" states of high emotional and motivational arousal greatly increase the accessibility of thoughts that relate to the immediate emotion and current needs, and reduce the accessibility of other thoughts (George Loewenstein, 1996).

Some attributes, which Tversky and Kahneman (1983) called *natural assessments*, are routinely and automatically registered by the perceptual system or by System 1, without intention or effort. Kahneman and Frederick (2002) compiled a list of natural assessments, with no claim to completeness. In addition to physical properties such as size, distance and loudness, the list includes more abstract properties such as similarity (e.g., Tversky & Kahneman, 1983), causal propensity (Kahneman & Varey, 1990; Heider, 1944; Michotte, 1963), surprisingness (Kahneman & Miller, 1986), affective valence (e.g., Bargh, 1997; Cacioppo, Priester, & Berntson, 1993; Kahneman, Ritov, & Schkade, 1999; Slovic, Finucane, Peters, & MacGregor, 2002; Zajonc, 1980), and mood (Schwarz & Clore, 1983). Accessibility itself is a natural assessment – the routine evaluation of cognitive fluency in perception and memory (e.g., Jacoby & Dallas, 1981; Johnson, Dark, & Jacoby, 1985; Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973).[1]

[1] The availability heuristic is based on an assessment of accessibility, in which frequencies or probabilities are judged by the ease with which instances come to mind. Tversky and I were responsible for this terminological confusion (Tversky and Kahneman, 1973).

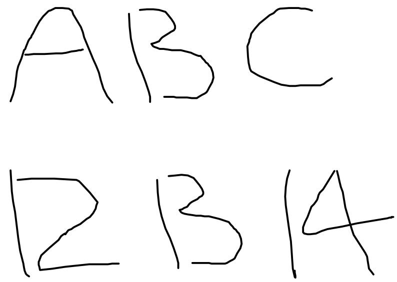


Figure 4.

Figure 4 illustrates the effect of context on accessibility. An ambiguous stimulus that is perceived as a letter in a context of letters is seen as a number in a context of numbers. The figure also illustrates another point: the ambiguity is suppressed in perception. This aspect of the demonstration is spoiled for the reader who sees the two versions in close proximity, but when the two lines are shown separately, observers will not spontaneously become aware of the alternative interpretation. They 'see' the interpretation that is the most likely in its context, but have no subjective indication that it could be seen differently. Similarly, in bi-stable pictures such as the mother/daughter figure or the Necker cube, there is no perceptual representation of the instability. And almost no one (for a report of a tantalizing exception, see Wittreich, 1961) is able to see the Ames room as anything but rectangular, even when fully informed that the room is distorted, and that the photograph does not provide enough information to specify its true shape. As the transactionalists who built the Ames room emphasized, perception is a choice of which we are not aware, and we perceive what has been chosen.

The unpredictability that is perceived as inherent to some causal systems is psychologically distinct from epistemic uncertainty, which is attributed to one's own ignorance (Kahneman & Tversky, 1982b). Competing propensities are often perceived – as they are when we watch a close horse race. And counterfactual alternatives to what happened are also perceived – we can see a horse that was catching up at the finish as 'almost winning the race' (Kahneman & Varey, 1990). In contrast to competing propensities, however, competing interpretations of reality appear to suppress each other: we do not see each horse in a close finish as both winning and losing. Epistemic uncertainty and ambiguity are not natural assessments.

Uncertainty is poorly represented in intuition, as well as in perception. Indeed, the concept of judgment heuristics was invented to accommodate the observation that intuitive judgments of probability are mediated by attributes such as similarity and associative fluency, which are not intrinsically related to uncertainty. The central finding in studies of intuitive decisions, as described by Klein (1998), is that experienced decision makers working under pressure, such as captains of firefighting companies, rarely need to choose between options because in most cases only a single option comes to their mind. The options that were rejected are not represented. Doubt is a phenomenon of

System 2, a meta-cognitive appreciation of one's ability to think incompatible thoughts about the same thing.

As this discussion illustrates, much is known about the determinants of accessibility, but there is no general theoretical account of accessibility and no prospect of one emerging soon. In the context of research in judgment and decision making, however, the lack of a theory does little damage to the usefulness of the concept. For most purposes, what matters is that empirical generalizations about the determinants of accessibility are widely accepted – and, of course, that there are procedures for testing their validity. For example, the claims about differential accessibility of different attributes in Figures 2 and 3 appealed to the consensual judgments of perceivers, but claims about accessibility are also testable in other ways. In particular, judgments of relatively inaccessible properties are expected to be substantially slower and more susceptible to interference by concurrent mental activity. Some tasks can be performed even while retaining several digits in memory for subsequent recall, but the performance of more effortful tasks will collapse under cognitive load.

Considerations of accessibility and analogies between intuition and perception play a central role in the programs of research that I will briefly review in what follows. Framing effects in decision making (Section 3) arise when different descriptions of the same problem highlight different aspects of the outcomes. The core idea of prospect theory (Section 4) is that changes and differences are much more accessible than absolute levels of stimulation. Judgment heuristics, which explain many systematic errors in beliefs and preferences are explained in Section 5 by a process of attribute substitution: people sometimes evaluate a difficult attribute by substituting a more accessible one. Variations in the ability of System 2 to correct or override intuitive judgments are explained by variations in the accessibility of the relevant rules (Section 6). Diverse manifestations of the differential accessibility of averages and sums are discussed in Section 7.

2. FRAMING EFFECTS

In Figure 2, the same property (the total height of a set of blocks) is highly accessible in one display and not in another, although both displays contain the same information. This observation is entirely unremarkable – it does not seem shocking that some attributes of a stimulus are automatically perceived while others must be computed, or that the same attribute is perceived in one display of an object but must be computed in another. In the context of decision making, however, similar observations raise a significant challenge to the rational-agent model. The assumption that preferences are not affected by variations of irrelevant features of options or outcomes has been called extensionality (Arrow, 1982) and invariance (Tversky & Kahneman, 1986). Invariance is an essential aspect of rationality, which is violated in demonstrations of *framing effects* such as the Asian disease problem (Tversky & Kahneman, 1981):

Problem 1 – The Asian Disease

Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved

Which of the two programs would you favor?

In this version of the problem, a substantial majority of respondents favor program A, indicating risk aversion. Other respondents, selected at random, receive a question in which the same cover story is followed by a different description of the options:

If Program A' is adopted, 400 people will die

If Program B' is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die

A clear majority of respondents now favor program B', the risk-seeking option. Although there is no substantive difference between the versions, they evidently evoke different associations and evaluations. This is easiest to see in the certain option, because outcomes that are certain are over-weighted relative to outcomes of high or intermediate probability (Kahneman & Tversky, 1979). Thus, the certainty of saving people is disproportionately attractive, and the certainty of deaths is disproportionately aversive. These immediate affective responses respectively favor A over B and B' over A'. As in Figures 2a and 2b, the different representations of the outcomes highlight some features of the situation and mask others.

The question of how to determine whether two decision problems are 'the same' or different does not have a general answer. To avoid this issue, Tversky and I restricted framing effects to discrepancies between choice problems that decision makers, upon reflection, consider effectively identical. The Asian disease problem passes this test: respondents who are asked to compare the two versions almost always conclude that the same action should be taken in both. Observers agree that it would be frivolous to let a superficial detail of formulation determine a choice that has life and death consequences.

In another famous demonstration of an embarrassing framing effect, McNeill, Pauker, Sox and Tversky (1982) induced different choices between surgery and radiation therapy, by describing outcome statistics in terms of survival rates or mortality rates. Because 90% short-term survival is less threatening than 10% immediate mortality, the survival frame yielded a substantially higher preference for surgery. The framing effect was no less pronounced among experienced physicians than it was among patients.

Shafir (1993) presented respondents with problems in which they played the role of a judge in adjudicating the custody of a child between divorcing parents. Each parent was described by a list of attributes. One of the descriptions was richer than the other: it contained more negative and more positive attributes. The framing of the instruction was varied. Some respondents were asked which custody request should be accepted, others decided which request should be rejected. The rich description was favored under both instructions, presumably because the respondents attended to its many advantages in deciding which custody request to accept, and to its many disadvantages in deciding about rejection.

A large-scale study by LeBoeuf and Shafir (in press) examined an earlier claim that framing effects are reduced, in a between-subjects design, for participants with high scores on 'need for cognition' (Smith & Levin, 1996). The original effect was not replicated in the more extensive study. However, LeBoeuf, and Shafir (2003) showed that more thoughtful individuals do show greater consistency in a within-subject design, where each respondent encounters both versions of each problem. This result is consistent with the present analysis. Respondents characterized by an active System 2 are more likely than others to notice the relationship between the two versions and to ensure the consistency of the responses to them. Thoughtfulness confers no advantage in the absence of a relevant cue, and is therefore irrelevant to performance in the between-subjects design.

Framing effects are not restricted to decision-making: Simon and Hayes (1976) documented an analogous observation in the domain of problem solving. They constructed a collection of transformation puzzles, all formally identical to the tower of Hanoi problem, and found that these 'problem isomorphs' varied greatly in difficulty. For example, the initial state and the target state were described in two of the versions as three monsters holding balls of different colors. The state transitions were described in one version as changes in the color of the balls, and in the other as balls being passed from one monster to another. The puzzle was solved much more easily when framed in terms of motion. The authors commented that "It would be possible for a subject to seek that representation which is simplest, according to some criterion, or to translate all such problems into the same, canonical, representation…" but "subjects will not employ such alternative strategies, even though they are available, but will adopt the representation that constitutes the most straightforward translation…" (Simon & Hayes, 1976, p 183).

Passive adoption of the formulation given appears to be a general principle, which applies as well to these puzzles, to the displays of Figure 2, and to the standard framing effects. People do not spontaneously compute the height of a tower that could be built from an array of blocks, and they do not spontaneously transform the representation of puzzles or decision problems. It is of interest, however, that some specialized perceptual and cognitive systems exhibit a limited ability to generate canonical representations for particular types of stimuli. Having seen a face once from a particular angle, for example, observers will recognize it from another angle, and they will also identify a black and white picture of it, or even a contour drawing. But even the versatile face-recognition module has its limitations: its performance is quite poor in recognizing familiar faces that are shown upside down. The brain mechanisms that support the comprehension of language also have a substantial ability to strip the surface details and get to the gist of meaning in an utterance, but this ability is limited as well. Few of us are able to recognize '137 x 24' and '3,288' as 'the same' number without going through some elaborate computations. Invariance cannot be achieved by a finite mind.

The impossibility of invariance raises significant doubts about the descriptive realism of rational-choice models (Tversky & Kahneman, 1986). Absent a system that reliably generates appropriate canonical representations, intuitive decisions will be shaped by the factors that determine the accessibility of different features of the situation. Highly accessible features will influence decisions, while features of low accessibility will be largely ignored. Unfortunately, there is no reason to believe that the most accessible features are also the most relevant to a good decision.

3. CHANGES OR STATES: PROSPECT THEORY

A general property of perceptual systems is that they are designed to enhance the accessibility of changes and differences (Palmer, 1999). Perception is *reference-dependent*: the perceived attributes of a focal stimulus reflect the contrast between that stimulus and a context of prior and concurrent stimuli. Figure 5 illustrates reference dependence in vision. The two enclosed squares have the same luminance, but they do not appear equally bright. The point of the demonstration is that the brightness of an area is not a single-parameter function of the light energy that reaches the eye from that area. An account of perceived brightness also requires a parameter for a reference value (often called adaptation level), which is influenced by the luminance of neighboring areas.

Figure 5.

The reference value to which current stimulation is compared also reflects the history of adaptation to prior stimulation. A familiar demonstration involves three buckets of water of different temperatures, arranged from cold on the left to hot on the right, with tepid in the middle. In the adapting phase, the left and right hands are immersed in cold and hot water, respectively. The initially intense sensations of cold and heat gradually wane. When both hands are then immersed in the middle bucket, the experience is heat in the left hand and cold in the right hand.

Reference-dependence in choice

The facts of perceptual adaptation were in our minds when Tversky and I began our joint research on decision making under risk. Guided by the analogy of perception, we expected the evaluation of decision outcomes to be reference-dependent. We noted, however, that reference-dependence is incompatible with the standard interpretation of Expected Utility Theory, the prevailing theoretical model in this area. This deficiency can be traced to the brilliant essay that introduced the first version of expected utility theory (Bernoulli, 1738).

One of Bernoulli's aims was to formalize the intuition that it makes sense for the poor to buy insurance and for the rich to sell it. He argued that the increment of *utility* associated with an increment of wealth is inversely proportional to initial wealth, and from this plausible psychological assumption he derived that the utility function for wealth is logarithmic. He then proposed that a sensible decision rule for choices that involve risk is to maximize the expected utility of wealth (the moral expectation). This proposition accomplished what Bernoulli had set out to do: it explained risk aversion, as well as the different risk attitudes of the rich and of the poor. The theory of expected utility that he introduced is still the dominant model of risky choice. The language of Bernoulli's essay is prescriptive – it speaks of what is sensible or reasonable to do – but the theory is also intended to describe the choices of reasonable men (Gigerenzer *et al.,* 1989). As in most modern treatments of decision making, there is no acknowledgment of any tension between prescription and description in Bernoulli's essay. The idea that decision makers evaluate outcomes by the utility of final asset positions has been retained in economic analyses for almost 300 years. This is rather remarkable, because the idea is easily shown to be wrong; I call it Bernoulli's error.

Bernoulli's model is flawed because it is *reference-independent*: it assumes that the value that is assigned to a given state of wealth does not vary with the decision maker's initial state of wealth.[2] This assumption flies against a basic principle of perception, where the effective stimulus is not the new level of

[2] What varies with wealth in Bernoulli's theory is the response to a given *change* of wealth. This variation is represented by the curvature of the utility function for wealth. Such a function cannot be drawn if the utility of wealth is reference-dependent, because utility then depends not only on current wealth but also on the reference level of wealth.

stimulation, but the difference between it and the existing adaptation level. The analogy to perception suggests that the carriers of utility are likely to be gains and losses rather than states of wealth, and this suggestion is amply supported by the evidence of both experimental and observational studies of choice (see Kahneman & Tversky, 2000). The present discussion will rely on two thought experiments, of the kind that Tversky and I devised when we developed the model of risky choice that we called Prospect Theory (Kahneman & Tversky, 1979).

Problem 2

Would you accept this gamble?

50% chance to win $150
50% chance to lose $100

Would your choice change if your overall wealth were lower by $100?

There will be few takers of the gamble in Problem 2. The experimental evidence shows that most people will reject a gamble with even chances to win and lose, unless the possible win is at least twice the size of the possible loss (e.g., Tversky & Kahneman, 1992). The answer to the second question is, of course, negative.

Next consider Problem 3:

Problem 3

Which would you choose?

lose $100 with certainty

or

50% chance to win $50
50% chance to lose $200

Would your choice change if your overall wealth were higher by $100?

In Problem 3, the gamble appears much more attractive than the sure loss. Experimental results indicate that risk seeking preferences are held by a large majority of respondents in choices of this kind (Kahneman & Tversky, 1979). Here again, the idea that a change of $100 in total wealth would affect preferences cannot be taken seriously.

Problems 2 and 3 evoke sharply different preferences, but from a Bernoullian perspective the difference is a framing effect: when stated in terms of final wealth, the problems only differ in that all values are lower by $100 in Problem 3 – surely an inconsequential variation. Tversky and I examined many choice pairs of this type early in our explorations of risky choice, and concluded that the abrupt transition from risk aversion to risk seeking could not plausibly be explained by a utility function for wealth. Preferences appeared to be determined by attitudes to gains and losses, defined relative to a reference point, but Bernoulli's theory and its successors did not incorporate a reference point. We therefore proposed an alternative theory of risk, in which the carriers of utility are gains and losses – changes of wealth rather than states of wealth. Prospect theory (Kahneman & Tversky, 1979) embraces the idea that preferences are reference-dependent, and includes the extra parameter that is required by this assumption.

The distinctive predictions of prospect theory follow from the shape of the value function, which is shown in Figure 6. The value function is defined on gains and losses and is characterized by four features: (1) it is concave in the domain of gains, favoring risk aversion; (2) it is convex in the domain of losses, favoring risk seeking; (3) Most important, the function is sharply kinked at the reference point, and *loss-averse* – steeper for losses than for gains by a factor of about 2–2.5 (Kahneman, Knetsch, & Thaler, 1991; Tversky & Kahneman, 1992). (4) Several studies suggest that the functions in the two domains are fairly well approximated by power functions with similar exponents, both less than unity (Swalm, 1966; Tversky & Kahneman, 1992). However, the value function is not expected to describe preferences for losses that are large relative to total assets, where ruin or near-ruin is a possible outcome.

Bernoulli's error – the assumption that the carriers of utility are final states – is not restricted to decision making under risk. Indeed, the error of reference-independence is built into the standard representation of indifference maps. It is puzzling to a psychologist that these maps do not include a representation of the decision maker's current holdings of various goods – the counterpart of the reference point in prospect theory. The parameter is not included, of course, because consumer theory assumes that it does not matter.

The wealth frame

The idea that the carriers of utility are changes of wealth rather than asset positions was described as the cornerstone of prospect theory (Kahneman & Tversky, 1979, p. 273). This statement implied that choices are always made by considering gains and losses rather than final states, but that proposition

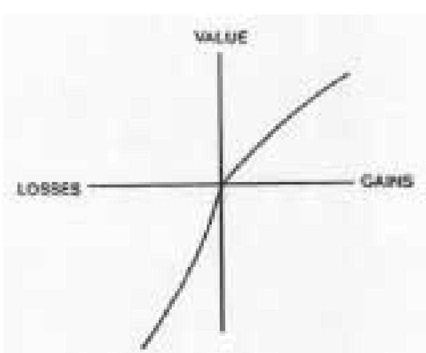


Figure 6.

must be qualified. The analysis of accessibility and framing that was presented earlier suggests a more moderate alternative, in which (1) decision problems can be formulated either in terms of wealth or in terms of changes; (2) the two formulations may lead to different preferences. For an example, consider Problem 4:

Problem 4

Please estimate your total wealth, call it W

Which of these situations is more attractive:

You own W

or

50% chance that you own W – $100
50% chance that you own W + $150

Informal experiments with problems of this type have consistently yielded a mild preference for the uncertain state of wealth, and a strong impression that the stakes mentioned in the question are entirely negligible.

In terms of final states of wealth, Problem 4 is identical to Problem 2. Furthermore, most respondents will agree, upon reflection, that the difference between the problems is inconsequential – too slight to justify different choices. Thus, the discrepant preferences observed in these two problems satisfy the definition of a framing effect.

The manipulation of accessibility that produces this framing effect is straightforward. The gamble of Problem 2 is likely to evoke an evaluation of the emotions associated with the immediate outcomes, and the formulation will not bring to mind thoughts of overall wealth. In contrast, the formulation of Problem 4 favors a view of the uncertainty as trivially small in relation to W, and includes no mention of gains or losses. In this perspective it is hardly surprising that the two problems elicit different representations, and therefore different preferences.

Over the centuries, Bernoulli's theory and its successors have been applied to decision problems in which outcomes are almost always formulated in terms of gains and losses, without any explicit mention of either current or final states of wealth. The assumption implicit in applications of expected utility theory is that outcomes described as gains or losses are first transformed into final asset states, then evaluated in that representation. In light of the preceding discussion of framing, the hypothesis of a transformation is highly implausible, and the different responses observed in Problems 2 and in Problem 4 provide direct evidence against it.

The same argument also applies in the other direction. Consider a decision maker who is only presented with Problem 4. Prospect theory assumed a preliminary operation of editing, in which prospects are reframed in simpler terms, prior to evaluation. But Problem 2 is not a simpler version of Problem 4; it includes gains and losses, which are not mentioned in Problem 4. The discussion of framing suggests that Problem 4 will be evaluated as it is stated – in terms of states of wealth. Indeed, some real-world choices are made in that frame. In particular, financial advisors and decision analysts often insist on formulating outcomes in terms of assets when they elicit their clients' preferences. Prospect theory is unlikely to provide an accurate description of decisions made in the wealth frame.

In experimental research as well as in the real world, the overwhelming majority of decisions are framed as gains and losses. There has been no systematic study of the choices that people make in the wealth frame, but one of the important properties of these choices is not in doubt: they will generally be closer to risk neutrality than when the equivalent outcomes are framed as gains and losses. The wealth frame favors risk neutrality in two ways. First, this frame eliminates any mention of losses, and therefore eliminates loss aversion. Second, in analogy with a familiar principle of perception, the outcomes of small bets will appear less significant when considered in the context of much larger amounts of wealth.

If Bernoulli's formulation is transparently incorrect as a descriptive model of risky choices, as has been argued here, why has this model been retained for so long? The answer may well be that the assignment of utility to wealth is an aspect of rationality, and therefore compatible with the general assumption of rationality in economic theorizing.

Consider Problem 5.

Problem 5

Two persons get their monthly report from a broker:

A is told that her wealth went from 4M to 3M
B is told that her wealth went from 1M to 1.1M

"Who of the two individuals has more reason
to be satisfied with her financial situation?"

"Who is happier today?"

Problem 5 highlights the contrasting interpretations of utility in theories that define outcomes as states or as changes. In Bernoulli's analysis only the first of the two questions is relevant, and only long-term consequences matter. Prospect theory, in contrast, is concerned with short-term outcomes, and the value function presumably reflects an anticipation of the valence and intensity of the emotions that will be experienced at moments of transition from one state to another (Kahneman, 2000a, b; Mellers, 2000). Which of these concepts of utility is more useful? For descriptive purposes, the more myopic notion is superior, but the prescriptive norms of reasonable decision making favor the long-term view. The Bernoullian definition of relevant outcomes is a good fit in a rational-agent model.

It is worth noting that an exclusive concern with the long term may be prescriptively sterile, because the long term is not where life is lived. Utility cannot be divorced from emotion, and emotion is triggered by changes. A theory of choice that completely ignores feelings such as the pain of losses and the regret of mistakes is not only descriptively unrealistic. It also leads to prescriptions that do not maximize the utility of outcomes as they are actually experienced – that is, utility as Bentham conceived it (Kahneman, 1994, 2000c; Kahneman, Wakker, & Sarin, 1997).

4. ATTRIBUTE SUBSTITUTION: A MODEL OF JUDGMENT BY HEURISTIC

The first joint research program that Tversky and I undertook was a study of various types of judgment about uncertain events, including numerical predictions and assessments of the probabilities of hypotheses. We reviewed this work in an integrative article (Tversky & Kahneman, 1974), which aimed to show "that people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors." (p. 1124). The second paragraph of that article introduced the idea that "the subjective assessment of probability resembles the subjective assessments of physical quantities such as distance or size. These judgments are all based on data of limited validity, which are processed according to heuristic rules." The concept of *heuristic* was illustrated by the role of the blur of contours as a potent determinant of the perceived distance of mountains. The observation that reliance on blur as a distance cue will cause distances to be overestimated on foggy days and underestimated on clear days was the example of a heuristic-induced *bias*. As this example illustrates, heuristics of judgment were to be identified by the characteristic errors that they inevitably cause.

Three heuristics of judgment, labeled representativeness, availability and anchoring, were described in the 1974 review, along with a dozen systematic biases, including non-regressive prediction, neglect of base-rate information, overconfidence and overestimates of the frequency of events that are easy to recall. Some of the biases were identified by systematic errors in estimates of known quantities and statistical facts. Other biases were identified by systematic discrepancies between the regularities of intuitive judgments and the principles of probability theory, Bayesian inference or regression analysis. The article defined the so-called "heuristics and biases approach" to the study of intuitive judgment, which has been the topic of a substantial research literature (Kahneman, Slovic, & Tversky, 1982; Gilovich, Griffin, & Kahneman, 2002) and has also been the focus of substantial controversy.

Shane Frederick and I recently revisited the conception of heuristics and biases, in the light of developments in the study of judgment and in the broader field of cognitive psychology in the intervening three decades (Kahneman & Frederick, 2002). The new model departs from the original formulation of heuristics in three significant ways: (i) it proposes a common process of attribute substitution to explain how judgment heuristics work; (ii) it extends the concept of heuristic beyond the domain of judgments about uncertain events; (iii) it includes an explicit treatment of the conditions under which intuitive judgments will be modified or overridden by the monitoring operations associated with System 2.

Attribute substitution

The 1974 article did not include a definition of judgmental heuristics. Heuristics were described at various times as principles, as processes, or as sources of cues for judgment. The vagueness did no damage, because the research program focused on a total of three heuristics of judgment under uncertainty, which were separately defined in adequate detail. In contrast, Kahneman and Frederick (2002) offered an explicit definition of a generic heuristic process of *attribute substitution*: A judgment is said to be mediated by a heuristic when the individual assesses a specified *target attribute* of a judgment object by substituting a related *heuristic attribute* that comes more readily to mind. This definition elaborates a theme of the early research, that people who are confronted with a difficult question sometimes answer an easier one instead. Thus, a person who is asked "What proportion of long-distance relationships break up within a year?" may answer as if she had been asked "Do instances of swift breakups of long-distance relationships come readily to mind?" This would be an application of the availability heuristic. A respondent asked to assess the probability that team A will beat team B in a basketball tournament may answer by mapping an impression of the relative strength of the two teams onto the probability scale (Tversky & Koehler, 1994). This could be called a "relative strength heuristic". In both cases, the target attribute is low in accessibility and another attribute, which is (i) related to the target, and (ii) highly accessible, is substituted in its place.

The word 'heuristic' is used in two senses in the new definition. The noun refers to the cognitive process, and the adjective in 'heuristic attribute' specifies the substitution that occurs in a particular judgment. For example, the representativeness heuristic is defined by the use of representativeness as a heuristic attribute to judge probability. The definition excludes anchoring effects, in which judgment is influenced by temporarily raising the accessibility

Figure 7.

of a particular *value* of the target attribute. On the other hand, the definition of the concept of heuristic by the process of attribute substitution greatly extends its range of application.

For a perceptual example of attribute substitution, consider the question: "What are the sizes of the two horses in Figure 7, as they are shown on the page?" The images are in fact identical in size, but the figure produces a compelling illusion. The target attribute that the observer is instructed to report is two-dimensional size, but the responses actually map an impression of three-dimensional size onto units of length that are appropriate to the required judgment. In the terms of the model, three-dimensional size is the heuristic attribute. As in other cases of attribute substitution, the illusion is caused by differential accessibility. An impression of three-dimensional size is the only impression of size that comes to mind for naïve observers – painters and experienced photographers are able to do better – and it produces a perceptual illusion in the judgment of picture size. The cognitive illusions that are produced by attribute substitution have the same character: an impression of one attribute is mapped onto the scale of another, and the judge is normally unaware of the substitution.

Direct tests of attribute substitution

An experiment described by Kahneman and Tversky (1973) illustrates a cognitive illusion that arises from attribute substitution. It also illustrates a particularly strict test of the hypothesis of substitution, in a research paradigm that Kahneman and Frederick (2002) labeled the *heuristic elicitation* design. Participants were given the following description of a fictitious graduate student, which was shown along with a list of nine fields of graduate specialization.

> Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense. (p.127)

Participants in a ***representativeness*** group ranked the nine fields of specialization by the degree to which Tom W. "resembles a typical graduate student" (in that field). Participants in a ***base-rate*** group evaluated the relative frequencies of the nine fields of graduate specialization. The description of Tom W. was deliberately constructed to make him more representative of the less populated fields: the rank correlation between the averages of representativeness rankings and of estimated base rates was -.65. Finally, participants in the ***probability*** group ranked the nine fields according to the likelihood of Tom W.'s specializing in each. These respondents were graduate students in psychology at major universities. They were given information that was intended to discredit the evidence of the personality sketch, namely that it had been written by a psychologist when Tom W. was in high school, on the basis of personality tests of dubious validity.

A description based on unreliable information should be given little weight, and predictions made in the absence of valid evidence should revert to base rates. Statistical logic therefore dictates that the correlation between judgments of probability and of representativeness should be negative in this problem. In contrast, the hypothesis of attribute substitution implies that the ranking of fields by the two measures should coincide. The results are shown in Figure 7. The correlation between the mean judgments of representativeness and of probability is nearly perfect (.97), supporting attribute substitution.

Another study in the same design involved one of the best-known characters in the heuristics and biases literature.

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.

Respondents were shown the description of Linda and a list of eight possible outcomes describing her present employment and activities. The two critical items in the list were #6 ("Linda is a bank teller") and the conjunction item #8 ("Linda is a bank teller and active in the feminist movement"). The other six possibilities were unrelated and miscellaneous (e.g., elementary school teacher, psychiatric social worker). As in the Tom W. problem, some respondents ranked the eight outcomes by representativeness; others ranked the same outcomes by probability. The correlation between the mean rankings was .99. Furthermore, the proportion of respondents who ranked the conjunction (item #8) higher than its constituent (#6) was about the same for representativeness (85%) and for probability (89%). The ordering of the two items is quite reasonable for judgments of similarity: Linda does resemble the

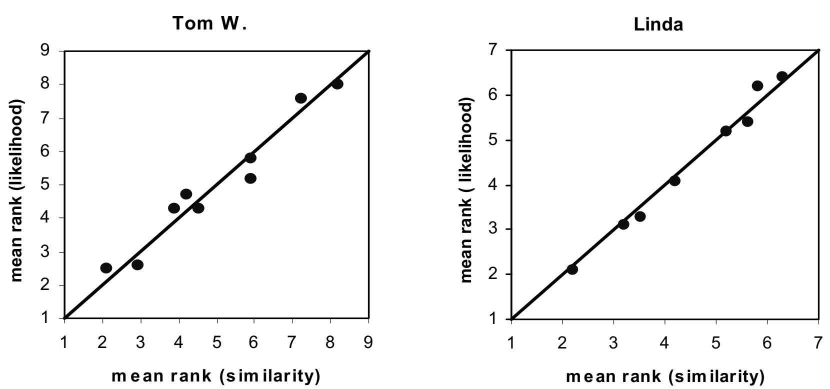


Figure 8.

image of a feminist bank teller more than she resembles a stereotypical bank teller. However, the reliance on representativeness as a heuristic attribute yields a pattern of probability judgments that violates monotonicity, and has been called the 'conjunction fallacy' (Tversky & Kahneman, 1983).

The results shown in Figure 8 are especially compelling because the responses were rankings. The large variability of the average rankings of both attributes indicates highly consensual responses, and nearly total overlap in the systematic variance. Stronger support for attribute-substitution could hardly be imagined, and it is surprising that this evidence was not acknowledged in subsequent debates about the validity of judgment heuristics. Other tests of representativeness in the heuristic elicitation design have been equally successful (Bar-Hillel & Neter, 2002; Tversky & Kahneman, 1982). The same design was also applied extensively in studies of support theory (Tversky & Koehler, 1994; for a review see Brenner, Koehler & Rottenstreich, 2002). In one of the studies reported by Tversky and Koehler (1994), participants rated the probability that the home team would win in each of 20 specified basketball games, and later provided ratings of the relative strength of the two teams, using a scale in which the strongest team in the tournament was assigned a score of 100. The correlation between normalized strength ratings and judged probabilities was .99.

The essence of attribute substitution is that respondents offer a reasonable answer to a question that they have not been asked. An alternative interpretation that must be considered is that the respondents' judgments reflect their understanding of the question they were asked. This may be true in some situations: it is not unreasonable to interpret a question about the probable outcome of a basketball game as referring to the relative strength of the competing teams. But the idea that judgments signify a commitment to the interpretation of the target attribute does not generally hold. For example, it is highly unlikely that educated respondents have a concept of probability that coincides precisely with similarity, or that they are unable to distinguish picture size from object size. A more plausible hypothesis is that an evaluation of the heuristic attribute comes immediately to mind, and that its associative relationship with the target attribute is sufficiently close to pass the monitoring of a permissive System 2. Respondents who substitute one attribute for another are not confused about the question that they are trying to answer – they simply fail to notice that they are answering a different one. And when they do notice the discrepancy, they either modify the intuitive judgment or abandon it altogether.

The new heuristics

As illustrated by its use in the interpretation of the visual illusion of Figure 7, the definition of judgment heuristics by the mechanism of attribute substitution applies to many situations in which people make a judgment that is not the one they intended to make. There is no finite list of heuristic attributes. Kahneman and Frederick (2002) illustrated this conception by a study by Strack, Martin, and Schwarz (1988), in which college students answered a survey that included these two questions: "How happy are you with your life in general?" and "How many dates did you have last month?". The correlation between the two questions was negligible when they occurred in the order shown, but it rose to 0.66 when the dating question was asked first. The model of attribute substitution suggests that the dating question automatically evokes an affectively charged evaluation of one's satisfaction in that domain of life, which lingers to become the heuristic attribute when the happiness question is subsequently encountered. The underlying correlation between the target and heuristic attributes is surely higher than the observed value of 0.66, which is attenuated by measurement error. The same experimental manipulation of question order was used in another study to induce the use of marital satisfaction as a heuristic attribute for well-being (Schwarz, Strack, & Mai, 1991). The success of these experiments suggests that *ad hoc* attribute substitution is a frequent occurrence.

The idea of an ***affect heuristic*** (Slovic *et al.,* 2002) is probably the most important development in the study of judgment heuristics in the last decades. There is compelling evidence for the proposition that every stimulus evokes an affective evaluation, which is not always conscious (see reviews by Zajonc, 1980, 1997; Bargh, 1997). Affective valence is a natural assessment, and therefore a candidate for substitution in the numerous responses that express attitudes. Slovic and his colleagues (Slovic *et al.,* 2002) discuss how a basic affective reaction can be used as the heuristic attribute for a wide variety of more complex evaluations, such as the cost/benefit ratio of technologies, the safe concentration of chemicals, and even the predicted economic performance of industries. Their treatment of the affect heuristic fits the present model of attribute substitution.

In the same vein, Kahneman and Ritov (1994) and Kahneman, Ritov, and Schkade (1999) proposed that an automatic affective valuation – the emotional core of an attitude – is the main determinant of many judgments and behaviors. In the study by Kahneman and Ritov (1994), 37 public causes were ranked by average responses to questions about (i) the importance of the issues, (ii) the size of the donation that respondents were willing to make, (iii) political support for interventions, and (iv) the moral satisfaction associated with a contribution. The rankings were all very similar. In the terms of the present analysis, the same heuristic attribute (affective valuation) was mapped onto the distinct scales of a wide range of target attributes. Similarly, Kahneman, Schkade, and Sunstein (1998) interpreted jurors' assessments of punitive awards as a mapping of outrage onto a dollar scale of punishments. In an article titled "Risk as Feelings", Loewenstein, Weber, Hsee, and Welch (2001), offered a closely related analysis in which emotional responses, such as the intensity of fear, govern diverse judgments (e.g., ratings of the probability of a disaster).

In terms of the scope of responses that it governs, the natural assessment of affect should join representativeness and availability in the list of general-purpose heuristic attributes. The failure to identify the affect heuristic much earlier, as well as its enthusiastic acceptance in recent years, reflect significant changes in the general climate of psychological opinion. It is worth noting that in the early 1970's the idea of purely cognitive biases appeared novel and distinctive, because the prevalence of motivated and emotional biases of judgment was taken for granted by the social psychologists of the time. There followed a period of intense emphasis on cognitive processes, in psychology generally and in the field of judgment in particular. It took another thirty years to achieve what now appears to be a more integrated view of the role of affect in intuitive judgment.

5. THE ACCESSIBILITY OF CORRECTIVE THOUGHTS

The present treatment assumes that System 2 is involved in all voluntary actions – including overt expressions of the intuitive judgments that originated in System 1. This assumption implies that errors of intuitive judgment involve failures of both systems: System 1, which generated the error, and System 2 which failed to detect and correct it (Kahneman & Tversky, 1982a). To illustrate this point, Kahneman and Frederick (2002) revisited the perceptual analogy that Tversky and Kahneman (1974) had used to explain how judgment heuristics generate biases: blur is a good cue to the distance of mountains, but reliance on this cue causes predictable errors of distance estimation on sunny or hazy days. The analogy was apt, but the analysis of the perceptual example neglected an important fact. Observers know, of course, whether the day is sunny or hazy, and they could use this knowledge to counteract the bias – but most often they do not. Contrary to what the early treatment implied, the use of blur as a cue does not inevitably lead to bias in the judgment of distance – the illusion could just as well be described as a failure to assign adequate negative weight to ambient haze. The effect of haziness on *impressions* of distance is a failing of System 1: the perceptual system is not designed to correct for this variable. The effect of haziness on *judgments* of distance is a separate failure of System 2. Analogous failures can be identified in other errors of intuitive judgment.

It is useful to consider how System 2 might have intervened in the problems of Tom W. and Linda that were described in an earlier section.

> "Tom W. does look like a library science person, but there are many more graduate students in Humanities and Social Sciences. I should adjust my rankings accordingly." "Linda cannot be more likely to be a feminist bank teller than to be a bank teller. I must rank these two outcomes accordingly"

These hypothetical samples of reasoning illustrate two ways in which intuitive judgments can be corrected. In the Tom W. example, the individual becomes aware of a factor that was not part of the intuitive judgment, and makes an effort to adjust accordingly. In the Linda example, the individual recognizes that the question can be answered by applying a decisive logical rule, which makes intuitions to the contrary irrelevant. Both would come under the rubric of "statistical heuristics", which people are sometimes capable of deploying in their reasoning about uncertain events (Nisbett, Krantz, Jepson, & Kunda, 1983/2002).

Neither of these examples of reasoning exceeds the intellectual reach of the graduate students at major universities whose rankings were shown in Figure 8. However, the data indicate that very few respondents actually came up with corrections. The puzzle is the same as in the blur illusion: why did these people not put their knowledge to good use? In the context of the present treatment, the question can be rephrased: Why did the statistical heuristics not become accessible when they were needed?

An important part of the answer is that attribute substitution is a silent process: the respondents who judge probability as if they had been asked to judge representativeness are not self-conscious about what they are doing. The substitute attribute is pertinent to the task, and its value comes to mind with little or no effort and with high confidence. There is therefore little reason for respondents to question their judgment, perhaps even less than in the bat-and-ball problem that was mentioned earlier. In contrast, the accessibility of statistical heuristics is often low, but it can be enhanced in at least two ways: by increasing the vigilance of the monitoring activities, or by providing stronger cues to the relevant rules.

A substantial research program was mounted by Nisbett, Krantz and their colleagues to investigate the factors that control the accessibility of statistical heuristics (Nisbett *et al.,* 1983/2002). For example, Nisbett *et al.* studied formally identical problems in several domains. They found that statistical reasoning was most likely to be evoked in the context of games of chance, occasionally evoked in situations involving sports, but relatively rare when the problems concerned the psychology of individuals. They also showed that the explicit mention of a sampling procedure facilitated statistical thinking (Nisbett *et al.,* 1983; see also Gigerenzer, Hell, & Blank, 1988). Zukier and Pepitone (1984) found that respondents were more likely to use base-rate information when instructed to think as statisticians than when instructed to emulate psychologists. Agnoli and Krantz (1989) found that brief training in the logic of sets improved performance in a simple version of the Linda problem. Considerations of accessibility are evidently relevant to the activation of statistical reasoning, not only to attribute substitution.

Nisbett, Krantz and their colleagues drew a sharp distinction between their statistical heuristics and the intuitive heuristics, which they described as "rapid and more or less automatic judgmental rules of thumb" (2002, p. 510). In the same vein, the present treatment assigns the competing heuristics to different cognitive systems. Attribute substitution has been described as an operation of System 1, which occurs automatically and effortlessly. In contrast, the statistical heuristics illustrate the rule-governed reasoning of System 2 (Sloman, 1996), which is deliberate and demands some effort. It is worth noting that the intervention of System 2 and the application of statistical heuristics and other rules do not guarantee a correct response. The rules that people apply in deliberate reasoning are sometimes false.

An implication of the view of intuition that has been proposed here is that statistical training does not eradicate intuitive heuristics such as representativeness, but only enables people to avoid some biases under favorable circumstances. The results shown in Figure 8, which were collected from statistically knowledgeable graduate students, support this prediction. In the absence of strong cues to remind them of their statistical knowledge, these respondents made categorical predictions like everybody else – by representativeness. However, statistical sophistication made a difference in a stripped-down version of the Linda problem, which required respondents to compare the probabilities of Linda being "a bank teller" or "a bank teller who is active in the feminist movement" (Tversky & Kahneman, 1983). The incidence of errors remained high for the statistically naïve even in that transparent version, but the error rate dropped dramatically among the sophisticated.

The efficacy of System 2 is impaired by time pressure (Finucane, Alhakami, Slovic, & Johnson, 2000) by concurrent involvement in a different cognitive task (Gilbert, 1989, 1991, 2002), by performing the task in the evening for 'morning people' and in the morning for 'evening people' (Bodenhausen, 1990), and, surprisingly, by being in a good mood (Isen, Nygren, & Ashby, 1988; Bless *et al.*, 1996). Conversely, the facility of System 2 is positively correlated with intelligence (Stanovich & West, 2002), with 'need for cognition' (Shafir & LeBoeuf, 2002), and with exposure to statistical thinking (Nisbett *et al.*, 1983; Agnoli & Krantz, 1989; Agnoli, 1991).

The observation that it is possible to design experiments in which 'cognitive illusions disappear' has sometimes been used as an argument against the usefulness of the notions of heuristics and biases (for example, Gigerenzer, 1991). In the present framework, however, there is no mystery about the conditions under which illusions appear or disappear. An intuitive judgment that violates a rule which the respondent accepts will be overridden, if the rule comes early enough to the respondent's mind. This argument is not circular, because we have adequate scientific knowledge (as well as widely shared folk knowledge) about the conditions that facilitate or impede the accessibility of logical or statistical rules.

The examples of possible corrections in the Tom W. and Linda problems illustrated two possible outcomes of the intervention of System 2: the intuitive judgment may be adjusted, or else rejected and replaced by another conclusion. A general prediction can be made about the former case, which is certainly the most frequent. Because the intuitive impression comes first, it is likely to serve as an anchor for subsequent adjustments, and corrective adjustments from anchors are normally insufficient. Variations on this theme are common in the literature (Epley & Gilovich, 2002; Epstein, 1994; Gilbert, 2002; Griffin & Tversky, 1992; Sloman, 2002; Wilson, Centerbar, & Brekke, 2002).

The methodological implication of this analysis is that intuitive judgments and preferences are best studied in between-subject designs. Within-subject designs with multiple trials encourage the adoption of simplifying strategies in which answers are computed mechanically, without delving into the specifics of each problem. Factorial designs are particularly undesirable, because they provide an unmistakable cue that every factor that is manipulated must be relevant to the judgment (Kahneman & Frederick, 2002). It is inappropriate to study intuitive judgments in conditions that are guaranteed to destroy their intuitive character. The difficulties of these experimental designs were noted long ago by Kahneman and Tversky (1982a), who pointed out that "Within-subject designs are associated with significant problems of interpretation in several areas of psychological research (Poulton, 1975). In studies of intuition, they are liable to induce the effect that they are intended to test" (p. 500). Unfortunately, this methodological caution has been widely ignored.

6. PROTOTYPE HEURISTICS

This section introduces a family of prototype heuristics, which share a common mechanism and a remarkably consistent pattern of cognitive illusions, analogous to the effects observed in the Tom W. and in the Linda problems (Kahneman & Frederick, 2002). Prototype heuristics can be roughly described as the substitution of an average for a sum – a process that has been extensively studied by Anderson in other contexts (e.g., Anderson, 1981, ch. pp. 58–70; 1991a,b). The section also discusses the conditions under which System 2 prevents or reduces the biases associated with these heuristics.

Extensional and prototype attributes

The target assessments in several significant tasks of judgment and decision making are *extensional attributes* of categories or sets. The value of an extensional attribute in a set is an aggregate (not necessarily additive) of the values over its extension. Each of the following tasks is illustrated by an example of an extensional attribute and by the relevant measure of extension. The argument of this section is that the extensional attributes in these tasks are low in accessibility, and are therefore candidates for heuristic substitution.

(i) category prediction (e.g., *the probability that the set of bank tellers contains Linda / the number of bank tellers*);
(ii) pricing a quantity of public or private goods (e.g., *the personal dollar value of saving a certain number of birds from drowning in oil ponds / the number of birds*);
(iii) global evaluation of a past experience that extended over time (e.g., *the overall aversiveness of a painful medical procedure / the duration of the procedure*);
(iv) assessment of the support that a sample of observations provides for a hypothesis (e.g., *the probability that a specified sample of colored balls has been drawn from one urn rather than another / the number of balls).*

Extensional attributes are governed by a general principle of conditional adding, which dictates that each element of the set adds to the overall value an amount that depends on the elements already included. In simple cases, the value is additive: the total length of the set of lines in Figure 3 is just the sum of their separate lengths. In other cases, each positive element of the set increases the aggregate value, but the combination rule is non-additive (typically sub-additive).[3]

A category or set which is sufficiently homogeneous to have a prototype can also be described by its *prototype attributes.* Where extensional attributes are akin to a sum, prototype attributes are averages. As the display of lines in Figure 3 illustrated, prototype attributes are often highly accessible. This observation is well-documented. Whenever we look at, or think about, an ensemble or category that has a prototype, information about the prototype becomes accessible. The classic discussion of basic-level categories included demonstrations of the ease with which features of the prototype come to mind (Rosch & Mervis, 1975). Even earlier, Posner and Keele (1968, 1970) had reported experiments in which observers were exposed on many trials to various distortions of a single shape. The prototype shape was never shown, but observers erroneously believed that it had been presented often. More recently, several studies in social psychology have shown that exposure to the name of a familiar social category increases the accessibility of the traits that are closely associated with its stereotype (see Fiske, 1998).

Because of their high accessibility, the prototype attributes are natural candidates for the role of heuristic attributes. A ***prototype heuristic*** is the label for the process of substituting an attribute of a prototype for an extensional attribute of its category (Kahneman & Frederick, 2002). The original instance of a prototype heuristic was the use of representativeness in category prediction. The probability of Linda being a bank teller is an extensional variable, but her resemblance to a typical bank teller is a prototype attribute.

Two tests of prototype heuristics

Because extensional and prototypical attributes are governed by characteristically different rules, the substitution of a prototype attribute for an extensional attribute entails two testable biases: extension neglect and violations of monotonicity. Tests of the two hypotheses are discussed in turn.

Tests of extension neglect

Doubling the frequencies of all values in a set will not affect prototype attributes, because measures of central tendency depend only on relative frequencies. In contrast, the value of an extensional attribute will increase monotonically with extension. The hypothesis that judgments of a target attribute are mediated by a prototype heuristic gains support if the judgments are insensitive to variations of extension.

The proposition that extension is neglected in a particular judgment has the character of a null hypothesis: it is strictly true only if all individuals in the

[3] If the judgment is monotonically related to an additive scale (such as the underlying count of the number of birds), the formal structure is known in the measurement literature as an "extensive structure" (Luce, Krantz, Suppes & Tversky, 1990, Chapter 3). There also may be attributes that lack any underlying additive scale, in which case the structure is known in the literature as a "positive concatenation structure" (Luce *et al.,* 1990, Chapter 19, vol. III, p. 38).

sample are completely insensitive to variations of extension. The hypothesis will be rejected, in a sufficiently large study, if even a small proportion of participants show some sensitivity to extension. The chances of some individuals responding to extension are high a priori, because educated respondents are generally aware of the relevance of this variable (Kahneman & Frederick, 2002). Everyone agrees that WTP for saving birds should increase with the number of birds saved, that extending a painful medical procedure by an extra period of pain makes it worse, and that evidence from larger samples is more reliable. Complete extension neglect is therefore an unreasonably strict test of prototype heuristics. Nevertheless, this extreme result can be obtained under favorable conditions, as the following examples show:

- The study of Tom W. (see Figure 8) illustrated a pattern of *base-rate neglect* in categorical prediction. This finding is robust when the task requires a ranking of multiple outcomes (Kahneman & Tversky, 1973). As noted in the preceding section, the sophisticated participants in this experiment were aware of the base-rates and were capable of using this knowledge in their predictions – but the thought of doing so apparently occurred to almost none of them. Kahneman and Tversky also documented almost complete neglect of base-rates in an experiment (the engineer/lawyer study) in which base-rates were explicitly stated. However, the neglect of explicit base-rate information in this design is a fragile finding (see Kahneman & Frederick, 2002; Koehler, 1996, Evans, Handley, Over, & Perham, 2002).
- Participants in a study by Desvousges *et al.,* (1993) indicated their willingness to contribute money to prevent the drowning of migratory birds. The number of birds that would be saved was varied for different sub-samples. The estimated amounts that households were willing to pay were $80, $78 and $88, to save 2,000, 20,000, or 200,000 birds, respectively. Frederick and Fischhoff (1998) reviewed numerous other demonstrations of *scope neglect* in studies of willingness to pay (WTP) for public goods. For example, Kahneman and Knetsch found that survey respondents in Toronto were willing to pay almost as much to clean up the lakes in a small region of Ontario or to clean up all the lakes in that province (reported by Kahneman, 1986).
- In a study described by Redelmeier and Kahneman (1996), patients undergoing colonoscopy reported the intensity of pain every 60 seconds during the procedure (see Figure 9), and subsequently provided a global evaluation of the pain they had suffered. The correlation of global evaluations with the duration of the procedure (which ranged from 4 to 66 minutes in that study) was .03. On the other hand global evaluations were correlated ($r = .67$) with an average of the pain reported on two occasions: when pain was at its peak, and just before the procedure ended. For example, patient A in Figure 9 reported a more negative evaluation of the procedure than patient B. The same pattern of *duration neglect* and Peak/End evaluations has been observed in other studies (Fredrickson & Kahneman, 1993; see Kahneman, 2000b, 2000c for a discussion).

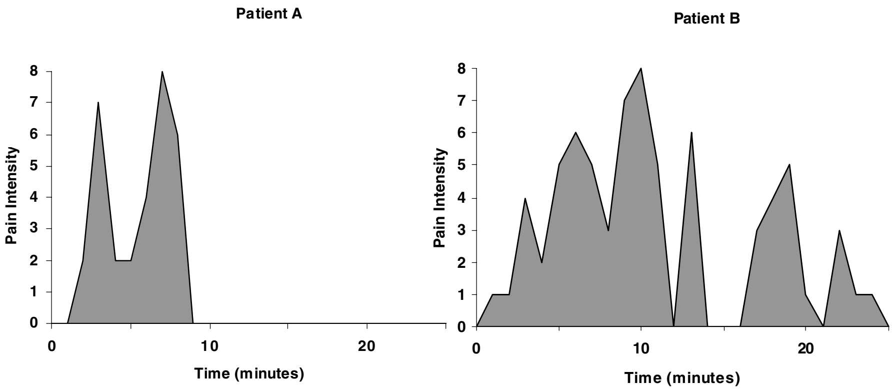


Figure 9. Pain intensity reported by two colonoscopy patients.

In light of the findings discussed in the preceding section, it is useful to consider situations in which people will *not* neglect extension completely. Extension effects are expected, in the present model, if the individual (i) has information about the extension of the relevant set; (ii) is reminded of the relevance of extension; and (iii) is able to detect that her intuitive judgments neglect extension. These conditions are least likely to hold – and complete neglect most likely to be observed – when the judge evaluates a single object and when the extension of the set is not explicitly mentioned. At the other extreme, the conditions for a positive effect of extension are all satisfied in psychologists' favorite research design: the within-subject factorial experiment, in which values of extension are crossed with the values of other variables in the design. As noted earlier, this design provides an obvious cue that the experimenter considers every manipulated variable relevant, and it enables participants to ensure that their judgments exhibit sensitivity to all these variables. The factorial design is therefore especially inappropriate for testing hypotheses about biases of neglect (Kahneman & Frederick, 2002).

In spite of these objections, within-subject factorial designs have been used in several experimental studies of extension neglect. Figure 10 illustrates the remarkably consistent *additive extension effect* that has emerged in these experiments (Schreiber & Kahneman, 2000). In each of the experiments, the extension variable has a slight but significant effect, and combines additively with other information. The additivity is noteworthy, because it is normatively inappropriate. For each panel of Figure 10, a compelling normative argument can be made for a quasi-multiplicative rule in which the lines should fan out.[4] The observed pattern is compatible with a process of anchoring and adjustment: the intuitive judgment provides an anchor, and small adjustments from that anchor are made to accommodate the role of extension.

[4] Anderson (1996, p. 253) has described several other situations in which variables that should be combined multiplicatively are combined additively.

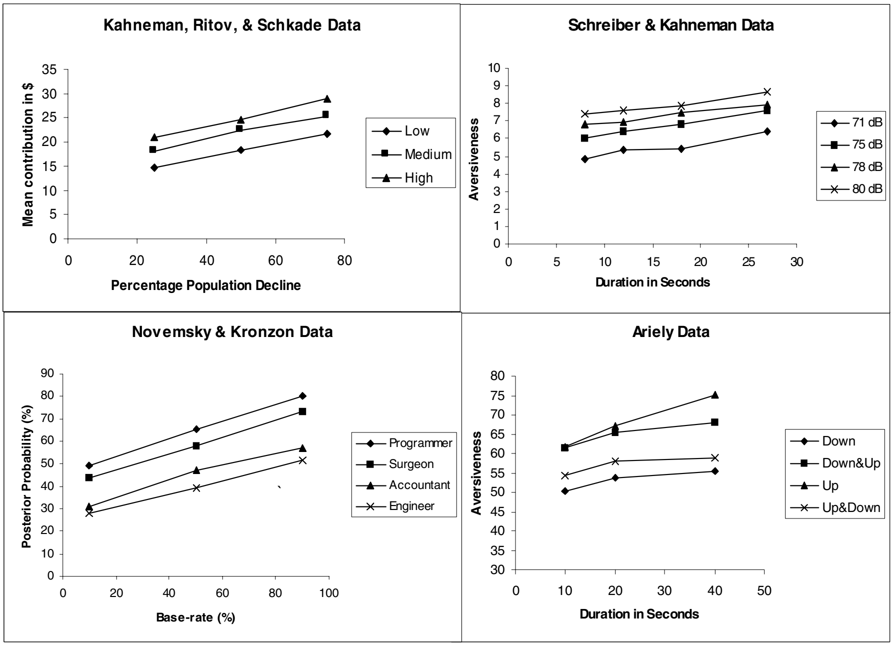


Figure 10. (a) Willingness to pay to restore damage to species that differ in popularity as a function of the damage they have suffered (from Kahneman, Ritov, & Schkade, 1999); (b) Global evaluations of aversive sounds of different loudness as a function of duration for subjects selected for their high sensitivity to duration (from Schreiber & Kahneman, 2000); (c) Ratings of probability for predictions that differ in representativeness as a function of base-rate frequency (from Novemsky & Kronzon, 1999); (d) Global evaluations of episodes of painful pressure that differ in temporal profile as a function of duration (Ariely, 1998).

Tests of monotonicity

Extensional variables, like sums, obey monotonicity. The sum of a set of positive values is at least as high as the maximum of its subsets. In contrast, the average of a subset can be higher than the average of a set that includes it. Violations of monotonicity are therefore bound to occur when an extensional attribute is judged by a prototype attribute: it must be possible to find cases in which adding elements to a set causes the judgment of the target variable to decrease. This test of prototype heuristics is less demanding than the hypothesis of extension neglect, and violations of monotonicity are compatible with some degree of sensitivity to extension (Ariely & Loewenstein, 2000). Nevertheless, violations of monotonicity in important tasks of judgment and choice are the strongest source of support for the hypothesis that prototype attributes are substituted for extensional attributes in these tasks.

- Conjunction errors, which violate monotonicity, have been demonstrated in the Linda problem and in other problems of the same type. There are no documented exceptions to the predicted pattern when the judgments are obtained in a between-subjects design, or when the two critical outcomes are embedded in a longer list (Tversky & Kahneman, 1982, 1983; Mellers, Hertwig, & Kahneman, 2001). Tversky and Kahneman (1983) also found that statistically naïve respondents made conjunction errors even in a direct comparison of the critical outcomes. As in the case of extension neglect, however, conjunction errors are less robust in within-subject conditions, especially when the task involves a direct comparison (see Kahneman & Frederick, 2002 for a discussion).

- Hsee (1999) asked participants to price sets of dinnerware offered in a clearance sale. One of the sets (A) consisted of 24 pieces, all in good condition. The other set included the same 24 pieces, plus 16 additional pieces, of which 7 were in a good condition and 9 were broken. When each respondent evaluated only one set, mean willingness to pay (WTP) was $33 for the smaller set and $23 for the larger set ($p < .01$). In contrast, participants who evaluated both sets were consistently willing to pay more for the larger set. List (2002) observed similar violations of dominance with a different good (sets of baseball cards), in a real market situation.
- Problems of the following kind have been used in several experiments (Kahneman & Tversky, 1972; Griffin & Tversky, 1992).

  A sample has been drawn from one of two urns. One urn contains 70% red balls and 30% white balls. The proportions are reversed in the other urn. What is the probability that each of these samples was drawn from the predominantly red urn?

  A sample of three red balls and zero white balls (3R, 0W)
  A sample of four red balls and three white balls (4R, 3W)
  A sample of seven red balls and three white balls (7R, 3W)

  The extensional target variable here is the degree of support for the 'red' hypothesis relative to the 'white' hypothesis. The normative solution is straightforward: posterior probability (the target attribute) is determined by an additive combination over sample elements – the difference between the number of red and white balls in the sample. The psychological solution is equally straightforward: the prototype attribute (the heuristic) is an average of support, which corresponds to the proportion of red balls in the sample. Thus, the addition of (4R, 3W) to (3R, 0W) raises the value of the target attribute but reduces the value of the heuristic attribute. This particular example is fictitious, but the pattern of findings indicates that respondents would derive much more confidence from (3R, 0W) than from (7R, 3W) (Kahneman & Tversky, 1972; Griffin & Tversky, 1992).
- A randomized clinical experiment was conducted as a follow-up to the colonoscopy study described earlier. For half the patients, the instrument was not immediately removed when the clinical examination ended. Instead, the physician waited for about a minute, leaving the instrument stationary. The experience during the extra period was uncomfortable, but the procedure guaranteed that the colonoscopy never ended in severe pain. Patients reported significantly more favorable global evaluations in this experimental condition than in the control condition (Redelmeier,

Katz, & Kahneman, in press). Violations of dominance have also been confirmed in choices. Kahneman, Fredrickson, Schreiber, and Redelmeier (1993) exposed participants to two cold-pressor experiences, one with each hand: a "short" episode (immersion of one hand in 14°C water for 60 seconds), and a "long" episode (the short episode, plus an additional 30 seconds during which the water was gradually warmed to 15°C). When they were later asked which of the two experiences they preferred to repeat, a substantial majority chose the long trial. This pattern of choices is predicted from the Peak/End rule of evaluation, which was described earlier. The same pattern of results was found with unpleasant sounds of variable loudness and duration (Schreiber & Kahneman, 2000).

The consistency of the results observed in diverse studies of prototype heuristics suggests the need for a unified interpretation, and challenges interpretations that only apply to a single domain. A number of authors have offered competing interpretations of base-rate neglect (Cosmides & Tooby, 1996; Koehler, 1996), insensitivity to scope in WTP (Kopp, 1992), and duration neglect (Ariely & Loewenstein, 2000). In general however, these interpretations are specific to a particular task, and would not carry over to demonstrations of extension neglect in the other tasks that have been discussed here. Similarly, the attempts to describe the conjunction fallacy as a miscommunication between experimenter and respondent (Dulany & Hilton, 1991; Hilton & Slugoski, 2001) do not explain analogous violations of monotonicity in the cold-pressor experiment and in the pricing of private goods. In contrast, the account offered here (and developed in greater detail by Kahneman & Frederick, 2002) is equally applicable to diverse tasks that require an assessment of an extensional target attribute.

The findings obtained in choices and joint evaluations confirm the existence of two distinct ways of choosing, which were already identified in prospect theory (Kahneman & Tversky, 1979). In the non-analytic procedure that I have called "choosing by liking" (Kahneman, 1994), the individual considers the global evaluation of the two options separately, and selects the one that has the higher global value, without detailed comparison of the options. Choice by global value was the basic mechanism assumed in prospect theory. However, prospect theory also introduced the idea that if the individual detects that one option dominates the other, the dominant option will be chosen without consulting their separate valuations. The same mechanisms apply to problems of judgment, such as the case of Linda, where some statistically sophisticated individuals detect that one of the sets includes the other and respond accordingly, ignoring representativeness. In Hsee's dinnerware study (1998), respondents chose by liking in separate evaluation, and chose by dominance in joint evaluation.

Joint evaluation is not sufficient to guarantee choice by dominance; it is also necessary for the decision makers to realize explicitly that one of the options is strictly better than the other. This requirement was not satisfied in the cold-pressor experiment. Although the participants were exposed to both experiences (joint evaluation), they did not notice that the long episode contained all the pain of the short one, and then some extra pain. Most respondents would have made a different choice if they had understood the structure of the options.

The normative logic of belief and choice is extensional, and it requires appropriate valuation of extensional attributes, which include both probability and utility. The examples that were discussed in this section demonstrate a pervasive departure from extensional logic, in the intuitive evaluation of both evidence and outcomes. The substitution of prototype attributes for extensional attributes appears to be a general characteristic of System 1, which is incompatible with both Bayesian beliefs and utility maximization.

CONCLUSIONS

The starting point of the present analysis was the observation that complex judgments and preferences are called 'intuitive' in everyday language if they come to mind quickly and effortlessly, like percepts. Another basic observation was that judgments and intentions are normally intuitive in this sense, but can be modified or overridden in a more deliberate mode of operation. The labels 'System 1' and 'System 2' were associated with these two modes of cognitive functioning.

The preceding sections elaborated a single generic proposition: "Highly accessible impressions produced by System 1 control judgments and preferences, unless modified or overridden by the deliberate operations of System 2." This template sets an agenda for research: to understand judgment and choice we must study the determinants of high accessibility, the conditions under which System 2 will override or correct System 1, and the rules of these corrective operations. Much is known about each of the three questions.

First, consider the ways in which the concept of accessibility was used here. Framing effects were attributed to the fact that alternative formulations of the same situation make different aspects of it accessible. The core idea of prospect theory, that the normal carriers of utility are gains and losses, invoked a general principle that changes are relatively more accessible than absolute values. Judgment heuristics were explained as the substitution of a highly accessible heuristic attribute for a less accessible target attribute. Finally, the proposition that averages are more accessible than sums unified the analysis of prototype heuristics. A recurrent theme was that different aspects of problems are made accessible in between-subjects and in within-subject experiments, and more specifically in separate and joint evaluations of stimuli. In all these cases, the discussion appealed to rules of accessibility that are independently plausible and sometimes quite obvious.

The status of accessibility factors in psychological theorizing is, in principle, similar to the status of perceptual grouping factors. In both cases there is no general theory, only a list of powerful empirical generalizations that provide a sound basis for experimental predictions and for models of higher-level phenomena. Unlike Gestalt principles, which were catalogued a long time ago, a comprehensive list of the factors that influence accessibility is yet to be drawn. The list will be long, but many of its elements are already known. For example, it is safe to assume that similarity is more accessible than probability, that changes are more accessible than absolute values, and that averages are more accessible than sums. Furthermore, each of these assumptions can be verified independently by multiple operations, including measurements of reaction time, susceptibility to interference by secondary tasks, and asymmetric priming. Assumptions about accessibility are incompletely theorized, but they need not be vague and they can do genuine explanatory work.

The present discussion of accessibility effects has been restricted to the differential accessibility of attributes (dimensions) on which judgment objects vary, such as length or price, similarity and probability, (Kahneman & Frederick, 2002). A similar analysis could be applied to the accessibility of particular values of attributes, such as 'six feet' or 'two dollars'. Highly accessible values are generally overweighted, and when considered as possible answers to a question they become potent anchors (Epley & Gilovich, 2002; Strack & Mussweiler, 1997; Chapman & Johnson, 2002). The effects of salience and anchoring play a central role in treatments of judgment and choice. Indeed, anchoring effects are among the most robust phenomena of judgment, and overweighting of salient values is likely to be the mechanism that explains why low-probability events sometimes loom large in decision making. The analysis of accessibility could readily be extended to deal with these observations.

The claim that cognitive illusions will occur unless they are prevented by System 2 sounds circular, but it is not. Circular inferences are avoidable because the role of System 2 can be independently verified in several ways. For example, the assumption that System 2 is vulnerable to interference by competing activities suggests that manifestations of intuitive thought that are normally inhibited may be expressed when people are placed under cognitive load. Another testable hypothesis is that intuitive judgments that are suppressed by System 2 still have detectable effects, e.g., in priming subsequent responses.

Principles of accessibility determine the relative power of the cues to which the monitoring functions of System 2 respond. For example, we know that differences between options are more salient in joint than in separate evaluation, and that any variable which is manipulated in a factorial design will attract some attention. Other cues can be found in the wording of problems and in the context of previous tasks. Many apparent inconsistencies in the literature on judgment heuristics are easily resolved within this framework (Kahneman & Frederick, 2002). A judgment bias that appears in some situations but not in others usually provides information about the factors that control corrective operations. As already noted, the attribution of the variability of intuitive judgments to System 2 is a source of readily testable hypotheses. It suggests, for example, that intelligence will be correlated with susceptibility to biases only in problems that provide relatively weak cues to the correct solution. In the absence of cues, there is no opportunity for intelligence or sophistication to manifest itself. When cues are abundant, at the other extreme, even the moderately intelligent will find them (Kahneman, 2000a; Stanovich & West, 1999, 2002).

The model suggests four ways in which a judgment or choice may be made:

(i) no intuitive response comes to mind, and the judgment is produced by System 2.
(ii) an intuitive judgment or intention is evoked, and
  a. is endorsed by System 2;
  b. serves as an anchor for adjustments that respond to other features of the situation;
  c. is identified as incompatible with a subjectively valid rule, and blocked from overt expression.

There is of course no way to ascertain precisely the relative frequencies of these outcomes, but casual observation suggests the following ordering, from most to least frequent:

(iia) – (iib) – (i) – (iic)

Most behavior is intuitive, skilled, unproblematic and successful (Klein, 1998). In some fraction of cases, a need to correct the intuitive judgments and preferences will be acknowledged, but the intuitive impression will be the anchor for the judgment. Under-correction is more likely than over-correction in such cases. A conservative general prediction is that variables that are neglected in intuition will remain underweighted in considered judgments.

The analysis of intuitive thinking and choice that has been presented here provides a framework which highlights commonalities between lines of research that are usually studied separately. In particular, the psychology of judgment and the psychology of choice share their basic principles, and differ mainly in content. At a more specific level, prototype heuristics solve structurally similar problems in diverse domains, where they yield closely similar patterns of results. Furthermore, the principles are not specific to the domain of judgment / decision making. The analogy between intuition and perception has been especially fruitful in identifying the ways in which intuitive thought differs from deliberate reasoning, and the notions of accessibility and dual-process analyses play a fundamental role in several domains of social and cognitive psychology.

A general framework such as the one offered here is not a substitute for domain-specific concepts and theories. For one thing, general frameworks and specific models make different ideas accessible. Novel ideas and compelling examples are perhaps more likely to arise from thinking about problems at a lower level of abstraction and generality. However, a broad framework can be useful if it guides a principled search for analogies across domains, to identify common processes and to prevent overly narrow interpretations of findings.

REFERENCES

Agnoli, F. (1991). Development of judgmental heuristics and logical reasoning: Training counteracts the representativeness heuristic. *Cognitive Development, 6*, 195–217.

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology, 21*, 515–550.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Anderson, N. H. (1991a). *Contributions to information integration theory* (Vol. I: Cognition). Hillsdale, NJ: Erlbaum.

Anderson, N. H. (1991b). *Contributions to information integration theory* (Vol. II: Social). Hillsdale, NJ: Erlbaum.

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

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

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157–162.

Ariely, D., & Loewenstein, G. (2000). When does duration matter in judgment and decision making? *Journal of Experimental Psychology: General, 129*, 524–529.

Arrow, K. J. (1982). Risk perception in psychology and economics. *Economic Inquiry, 20*, 1–9.

Bar-Hillel, M., & Neter, E. (2002). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases* (pp.82–97). New York: Cambridge University Press, 2002, 82–97.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *The automaticity of everyday life: Advances in social cognition* (Vol. 10, pp.1–61). Mahwah, NJ: Erlbaum.

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

Bless, H., Clore, G. L., Schwarz, N., Golisano, V., Rabe, C., & Wolk, M. (1996). Mood and the use of scripts: Does a happy mood really lead to mindlessness? *Journal of Personality and Social Psychology, 71*, 665–679.

Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science, 1*, 319–322.

Brenner, L. A., Koehler, D. J., & Rottenstreich, Y. (2002). Remarks on support theory: Recent advances and future directions. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases* (pp. 489–509). New York: Cambridge University Press.

Cacioppo, J. T., Priester, J. R., & 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.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology.* New York: Guilford Press.

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases* (pp.120–138). New York: Cambridge University Press.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research,* in press.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment and uncertainty. *Cognition, 58*, 1–73.

Desvousges, W. H., Johnson, F., Dunford, R., Hudson, S., Wilson, K., & Boyle, K. (1993). Measuring natural resource damages with contingent valuation: Tests of validity and reliability. In J. A. Hausman (Ed.), *Contingent valuation: A critical assessment* (pp. 91–159). Amsterdam: North Holland.

Dulany, D. E., & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition, 9*, 85–110.
Epley, N., & Gilovich, T. (2002). Putting adjustment back in the anchoring and adjustment heuristic. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases* (pp. 139–149). New York: Cambridge University Press.
Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49*, 709–724.
Evans, J. St. B. T., Handley, S. J., Over, D. E., & Perham, N. (2002). Background beliefs in Bayesian inference. *Memory and Cognition, 30*, 179–190.
Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*, 1–17.
Fiske, S. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 357–441). New York: McGraw-Hill.
Frederick, S. W., & Fischhoff, B. (1998). Scope (in)sensitivity in elicited valuations. *Risk, Decision, and Policy, 3*, 109–123.
Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology, 65*, 45–55.
Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." In W. Stroebe & M. Hawthorne (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 83–115). Chichester, UK: Wiley.
Gigerenzer, G., Hell, W. & Blank, H. (1988). Presentation and content – the use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 513–525.
Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krueger, L. (1989). *The empire of chance: How probability changed science and everyday life.* Cambridge, UK: Cambridge University Press.
Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). Englewood Cliffs, NJ: Prentice-Hall.
Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*, 107–119.
Gilbert, D. T. (2002). Inferential correction. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases* (pp.167–184). New York: Cambridge University Press.
Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and Biases.* New York: Cambridge University Press.
Griffin, D. W., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.
Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice.* New York: Oxford University Press.
Heider, F. (1944). Social perception and phenomenal casuality. *Psychological Review, 51*, 358–374.
Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp.133–168). New York: Guilford Press.
Hilton, D. J., & Slugoski, B. R. (2001). Conversational processes in reasoning and explanation. In A. Tesser & N. Schwartz (Eds.), *Blackwell handbook of social psychology.* (Vol. 1: Intraindividual processes, pp.181–206). Oxford, UK: Blackwell.
Hogarth, R. M. (2001). *Educating intuition.* Chicago: University of Chicago Press.
Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making, 11*, 107–121.
Hsee, C. K. (1999). Value seeking and prediction decision inconsistency: Why don't people take what they predict they'll like the most? *Psychonomic Bulletin and Review, 6*, 555–561.
Isen, A. M., Nygren, T. E., & Ashby, F. G. (1988). Influence of positive affect on the subjective utility of gains and losses: It is just not worth the risk. *Journal of Personality and Social Psychology, 55*, 710–717.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.
Jacoby, L. L. (1996). Dissociating automatic and consciously controlled effects of study/test compatibility. *Journal of Memory and Language, 35*, 32–52.
Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.
Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 3–11.
Kahneman, D. (1986). Comment. In, R. G. Cummings, D. S. Brookshire & W. D. Schultze (Eds.), *Valuing Environmental Goods* (pp.185–193). Totowa, NJ: Rowman & Allenheld.
Kahneman, D. (1994). New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics, 150*, 18–36.
Kahneman, D. (2000a). A psychological point of view: Violations of rational rules as a diagnostic of mental processes (Commentary on Stanovich and West). *Behavioral and Brain Sciences, 23*, 681–683.
Kahneman, D. (2000b). Experienced utility and objective happiness: A moment-based approach. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames* (pp.673–692). New York: Cambridge University Press.
Kahneman, D. (2000c). Evaluation by moments: Past and future. In D. Kahneman and A. Tversky (Eds.), *Choices, values, and frames* (pp.693–708). New York: Cambridge University Press.
Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases* (pp. 49–81). New York: Cambridge University Press.
Kahneman, D., Fredrickson, D. 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., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.
Kahneman, D., & Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study in the headline method. *Journal of Risk and Uncertainty, 9*, 5–38.
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, 19*, 220–242.
Kahneman, D., Schkade, D. A., & Sunstein, C. R. (1998). Shared outrage and erratic awards: The psychology of punitive damages. *Journal of Risk and Uncertainty, 16*, 49–86.
Kahneman, D., Slovic, P. & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.
Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430–454.
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., & Tversky, A. (1982a). On the study of statistical intuitions. In D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment under uncertainty: Heuristics and biases* (pp.493–508). New York: Cambridge University Press.
Kahneman, D., & Tversky, A. (1982b). Variants of uncertainty. *Cognition, 11*, 143–157.
Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames.* New York: Cambridge University Press.
Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won, *Journal of Personality and Social Psychology, 59*, 1101–1110.
Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics, 112*, 375–405.

Klein, G. (1998). *Sources of power: How people make decisions.* Cambridge: MIT Press.

Koehler, D. J. (1996). A strength model of probability judgments for tournaments. *Organizational Behavior and Human Decision Making Processes, 66,* 16–21.

Kopp, R. (1992). Why existence value should be used in cost-benefit analysis. *Journal of Policy Analysis and Management, 11,* 123–130.

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making,* in press.

List, J. (2002). Preference reversals of a different kind: The more is less phenomenon. *American Economic Review, 92,* 1636–1643.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes, 65,* 272–292.

Loewenstein, G., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127,* 267–286.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky A. (1990). *Foundations of measurement* (Vol. 3: Representation, axiomatization, and invariance). San Diego, CA: Academic Press.

McNeil, B. J., Pauker. S. G., Sox, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine, 306,* 1259–1262.

Mellers, B. (2000). Choice and the relative pleasure of consequences. *Psychological Bulletin, 126,* 910–924.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12,* 269–275.

Michotte, A. (1963). *The perception of causality* (T. R. Miles & E. Miles, Trans.). New York: Basic Books.

Myers, D. G. (2002). *Intuition: Its powers and perils.* New Haven, CT: Yale University Press.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (2002). The use of statistical heuristics in everyday inductive reasoning. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases* (pp.510–533). New York: Cambridge University Press.

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.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology.* Cambridge, MA: The MIT Press.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.

Poulton, E. C. (1975). Range effects in experiments with people. *American Journal of Psychology, 77,* 353–363.

Redelmeier, D., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain, 66,* 3–8.

Redelmeier, D. A., Katz, J., & Kahneman, D. (in press). Memories of colonoscopy: A randomized trial. *Pain.*

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 104,* 406–415.

Schreiber, C. A., & Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General, 129,* 27–42.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523.

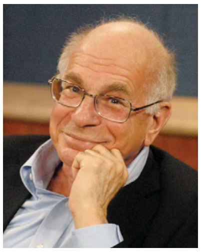
Schwarz, N., Strack, F., & Mai, H. P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly, 55*, 3–23.
Schwarz, N., & Vaughn, L. A. (2002). The availability heuristic revisited: Ease of recall and content of recall as distinct sources of information. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases* (pp. 103–119). Cambridge: Cambridge University Press.
Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition, 21*, 546–556.
Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology, 53*, 419–517.
Simon, H. A., & Hayes, J. R. (1976). Understanding process: Problem isomorphs. *Cognitive Psychology, 8*, 165–190.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22.
Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases* (pp.379–396). Cambridge: Cambridge University Press.
Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases* (pp.397–420). Cambridge: Cambridge University Press.
Smith, S. M., & Levin, I. P. (1996). Need for cognition and choice framing effects. *Journal of Behavioral Decision Making, 9*, 283–290.
Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.
Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding / acceptance principle. *Cognitive Psychology, 38*, 349–385.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23*, 645–665.
Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases* (pp.421–440). Cambridge: Cambridge University Press.
Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life-satisfaction. *European Journal of Social Psychology, 18*, 429–442.
Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*, 437–446.
Swalm, R. O. (1966). Utility theory: Insights into risk taking. *Harvard Business Review, 44*, 123–136.
Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105–110.
Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability, *Cognitive Psychology, 5*, 207–232.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp.153–160). New York: Cambridge University Press.
Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.
Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59*, S251–0S278.
Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics, 106*, 1039–1061.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty, 5*, 297–323.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*, 547–567.
Wilson, T. D., Centerbar, D. B., & Brekke, N. (2002). Mental contamination and the debiasing problem. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases* (pp.185–200). Cambridge: Cambridge University Press.
Wittreich, W. J. (1961). The Honi phenomenon: A case of selective perceptual distortion. In F. P. Kilpatrick (Ed.), *Explorations in transactional psychology* (pp. 188–202). New York: New York University Press.
Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151–175.
Zajonc, R. B. (1997). Emotions. In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th Ed., Vol. 1, pp. 591–632). New York: Oxford University Press.
Zukier, H., & Pepitone, A. (1984). Social roles and strategies in prediction: Some determinants in the use of base-rate information. *Journal of Personality and Social Psychology, 47*, 349–360.

The Human Side of Decision Making

Thinking Things Through with Daniel Kahneman, PhD

Daniel Kahneman is widely considered the most influential psychologist in the world today. He is best known in the financial realm for pioneering work that helped to lay the foundation for behavioral economics, which studies the psychology of judgment and economic decision making and its impact on the financial markets. Together with his long-time collaborator Amos Tversky, Dr. Kahneman explored the ways in which human judgment systematically departs from the basic principles of decision theory when evaluating economic risk, consequently creating the concept of prospect theory. Their findings challenged fundamental economic assumptions and expanded the boundaries of research by introducing psychologically realistic models into economic theory. So ground-breaking are their discoveries that *New York Times* columnist David Brooks has called Drs. Kahneman and Tversky "the Lewis and Clark of the mind." In 2002, Dr. Kahneman's work was recognized with the Nobel Memorial Prize in Economic Sciences for his integration of insights from psychological research into economic science.



Daniel Kahneman, PhD

Born in Tel Aviv in 1934, Dr. Kahneman spent his childhood—including the period of the Nazi occupation (1940–1944)—in France before moving to British Palestine (now Israel) in 1948. In 1954, he earned a bachelor of science degree with a major in psychology and a minor in mathematics from The Hebrew University of Jerusalem, then joined the Israel Defense Forces, where he served in the psychology branch. His responsibilities included evaluating candidates for officer training school and developing a method for interviewing combat unit recruits, which much later provided some of the basic ideas of his work with Amos Tversky on judgment. According to Dr. Kahneman, "This was the beginning of a lifelong interest in the statistics of prediction and description." In 1958, he began PhD studies at the University of California, Berkeley. After completing a doctorate in psychology in 1961, Dr. Kahneman returned to The Hebrew University of Jerusalem as a lecturer in psychology; he was promoted to senior lecturer in 1966 and later to professor.

In 1969, Dr. Kahneman began his long collaboration with Dr. Tversky, a fellow psychology professor at The Hebrew University. Their first jointly authored paper, "Belief in the Law of Small Numbers," was published in 1971. Over the next thirteen years, Dr. Kahneman and Dr. Tversky worked together to produce a series of seminal articles in the field of judgment and decision making. Among the most important of these were "Judgment under Uncertainty: Heuristics and Biases," published in 1974 in *Science*, which introduced the idea of judgment heuristics, including anchoring; and "Prospect Theory: An Analysis of Decision under Risk," published in 1979 in *Econometrica*. In 1977, Dr. Kahneman and Dr. Tversky met Richard Thaler, who later became the leading figure in behavioral economics. Dr. Kahneman has called his friendship with Dr. Thaler "the second most important professional friendship" of his life. Dr. Kahneman and Dr. Tversky subsequently became involved in the development of this new approach to economic theory, eventually collaborating on several papers with Dr. Thaler.

In 1978, Dr. Kahneman moved to Vancouver to take a position as professor of psychology at the University of British Columbia. He continued to collaborate with Dr. Tversky, who had accepted a position at Stanford University the same year, and the two completed their study of framing over the next several years. Dr. Kahneman also collaborated with Dr. Thaler on a variety of topics that integrated psychology and economics, including the endowment effect and public views about fairness in economic transactions. From 1986 to 1993, Dr. Kahneman returned to the University of California, Berkeley, as professor of psychology. During the 1990s, Dr. Kahneman's research focus shifted to hedonic psychology—the study of what makes experiences and life pleasant or unpleasant, satisfying or unsatisfying—as well as to studies of well-being that built on his previous research about experienced utility. Recently he has been working to develop and promote adversarial collaboration within the social sciences. During the course of his academic career, Dr. Kahneman also has been associated with the University of Michigan, Harvard University, the Russell Sage Foundation, the Canadian Institute for Advanced Research, and the Applied Psychological Research Unit in Cambridge, England.

Since 1993, Dr. Kahneman has been associated with Princeton University, where he is the Eugene Higgins Professor of Psychology, Emeritus, and Professor of Psychology and Public Affairs, Emeritus; he is also a Senior Scholar at Princeton's Woodrow Wilson School of Public and International Affairs and a Fellow at the Center for Rationality at The Hebrew University of Jerusalem. Since 2004, he has served as a Gallup senior scientist, advising and consulting with Gallup researchers on behavioral economics and his recent research on psychological well-being. Dr. Kahneman is a founding partner of The Greatest Good, a business and philanthropy consulting company formed with the goal of applying cutting-edge data analysis and economic methods to the most salient problems in business. He is a consultant to Guggenheim Partners, an investment advisory firm.

Dr. Kahneman has written and edited numerous books and authored more than 170 articles for professional journals. The 1974 *Science* paper and the 1979 *Econometrica* paper that he co-authored with Dr. Tversky are among the most frequently quoted works in social science; Dr. Kahneman himself was cited in scholarly journals more than 28,000 times between 1979 and 2011, according to the Thomson Reuters Web of Science data base. The Decision Analysis Society presented Dr. Kahneman with its Publication Award for the best paper published in 2003 for "Maps of Bounded Rationality: Psychology for Behavioral Economics." Dr. Kahneman's recent book, *Thinking, Fast and Slow* (2011), summarizes much of his research, is a bestseller, was selected as one of the best books of 2011 by the *New York Times Book Review*, the *Wall Street Journal*, the *Economist*, and Canada's *Globe and Mail*; and won the *Los Angeles Times* Book Prize.

> "As you know, most of my research has been collaborative. So having brilliant friends, I think, is the secret of any success I have achieved."

Dr. Kahneman has received every major award in the field of psychology, including the American Psychological Association's 2007 award for outstanding lifetime contributions to psychology and the University of Louisville's 2003 Grawemeyer Award in psychology (with Dr. Tversky) for revolutionizing the scientific study of decision making. In 2011, the American Economic Association named Dr. Kahneman as a distinguished fellow, *Bloomberg* named him as one of the fifty most influential people in global finance, and *Foreign Policy* magazine recognized him as one of the world's top global thinkers. In 2005, Dr. Kahneman was voted 101st among the 200 greatest Israelis of all time in a poll conducted by the Israeli news website Ynet. He has been awarded honorary doctorates by numerous universities including the University of Michigan, Erasmus University in the Netherlands, University of Paris, University of Milan, Harvard University, and The New School.

In February 2012, Dr. Kahneman spoke with members of the Editorial Advisory Board of the *Journal of Investment Consulting* about his investigations into decision making in the context of a dual-process model, loss aversion and risk tolerance, adversarial collaboration, and financial advisors' impact on investors' well-being. Taking part in the discussion were Margaret M. Towle, PhD, the *Journal* editor-in-chief, of HighTower Advisors; Mark Anson, PhD, of Oak Hill Investments; Edward Baker of The Cambridge Strategy; Geoffrey Gerber, PhD, of TWIN Capital Management; and Meir Statman, PhD, of Santa Clara University. This interview is the twelfth in the *Journal*'s Masters Series, which presents topical discussions with leading experts and visionaries in finance, economics, and investments.

Margaret Towle: First of all, Dr. Kahneman, thank you so much for agreeing to spend some time with us today. We're all well-acquainted with your exceptional background and contributions, and we hope to get a little more insight into the factors that helped to shape your career. Looking back over your experiences—from your childhood in Nazi-occupied France, your collaboration with Richard Thaler[1] in behavioral economics, through your recent work in intuition and the role that it plays in scientific investigation—what do you regard as the major factors that shaped your career and brought you to where you are today? Your accomplishments are too numerous to list, of course, but what do you consider your major achievements?

Daniel Kahneman: I think there's no question about the main determinant of my career, and that was the joint work with Amos Tversky.[2] As you know, most of my research has been collaborative. So having brilliant friends, I think, is the secret of any success I have achieved. In addition, there is a large element of being in the right place at the right time intellectually, that is, answering questions to which people are interested in hearing your answers. So, yes, I've been very fortunate. Clearly, if you want to understand what I've done, it's mostly collaborative.

Margaret Towle: What about on the other side, that is, what you would consider—I don't know if we want to call them mistakes—but your biggest disappointments in terms of events that happened throughout your career?

Daniel Kahneman: The worst thing that happened in my career was that, as I just mentioned, Amos Tversky and I collaborated for a long time, beginning in 1969, and then [in 1978] he went to Stanford University and I went to the University of British Columbia. We went on collaborating for a while after that, but it became very difficult for many reasons, mainly the physical separation. I think that together we were doing work that was better than either of us did separately. So the fact that we stopped working together was a major disappointment. I think I would have done better work if we had gone on working together, and probably so would he.

Meir Statman: In your most recent book, *Thinking, Fast and Slow*, you talk about the organizing principles of System 1 and System 2.[3] I was speaking some months ago to a group of wealthy investors and business owners, noting the need to check intuition by the rules of science. One of the participants said that he still trusts his gut much more than scientific evidence. How can we persuade people to check their intuition? And should we persuade people to check their intuition?

Daniel Kahneman: I don't know that you can persuade everybody. The confidence that people have in their intuitions is a genuine feeling; it is not an opinion. You have the immediate feeling that your thinking is right, that your intuitions are valid, and it's like something you see, an illusion. People are very resistant to changing their minds about their cognitive illusions. We're much more willing to accept visual illusions, but people really resist when you tell them that their thinking in a certain way is an illusion. It's very difficult to convince them. On the whole, the ideas of System 1 and System 2 are penetrating, that is, there is more and more readiness to accept them. However, it's slow, and when they conflict with people's direct intuition, you'll find they quite frequently lose.

Meir Statman: The people to whom I was speaking were members of families who had established very successful businesses. I was wondering whether their experience had involved one or two decisions that went spectacularly well, which persuaded them to believe in a version of the law of small numbers.[4]

Daniel Kahneman: Absolutely. It's very clear that it doesn't take very much for people to think that there is a pattern, and it doesn't take many successes for people to think that they are very, very smart, and it doesn't take many successes for others to think that a successful person has been very smart. People can be lucky, and that will feed into overconfidence. But even without luck, people are prone to overconfidence.

Ed Baker: I have a slightly different question, but related to that. I picked up on one comment you made in your Nobel Prize autobiography, which I found to be just fascinating. In particular, you said that most highly cognitive performances are intuitive. I wondered, when it comes to identifying skill, does that make it harder or easier? Is there something about this characteristic that one can identify, or is it really just unique from instance to instance? Is there a pattern that one can see?

Daniel Kahneman: What we call intuitive thinking refers to the ideas that come to mind quickly and without reflection, quite often automatically. You're in a situation, and you know what to do or you know how to understand that situation. Most of the time, our intuitions are just fine. We mostly run on what I call System 1 intuitively and with high confidence and very successfully. That is true both in very simple matters—for example, recognizing a speaker's emotion on the telephone from hearing one word, this is something at which all of us are quite skilled. Intuition is often excellent in complex tasks as well. We have learned hundreds of skills that actually are at the level of a chess master, except we don't think of them that way. When we get highly practiced, we develop skills. The problem with intuition and with people who want to trust their gut is that intuitions come with high confidence. The confidence is justified when intuition is a product of skill, which people have acquired through numerous experiences with immediate feedback. However, some intuitions are products of heuristics of judgment and are quite often mistaken.[5] The problem is that even the mistaken intuitions come to mind with considerable confidence. It's very difficult to distinguish between intuitions that reflect real skill and intuitions that don't. It is not easy for an observer, and even harder for the individual who has the intuition. We don't know the boundary between skill and heuristic in our own thinking.

Ed Baker: So that makes this type of behavior very difficult to distinguish, that is, when it's an example of skill and when it's not?

> "Optimism also facilitates resiliency in the context of execution. However, we need to distinguish situations in which optimism and confidence are useful from situations in which they are not."

Daniel Kahneman: In *Thinking, Fast and Slow*, I described my collaborative work with Gary Klein[6] on determining whether you can trust intuitive thinking. The conclusion is that if you want to know whether you can trust intuition, your own or somebody else's, you shouldn't ask about subjective confidence, because that can be very misleading. Instead, you should ask about the probability that a person's intuitions arise from genuine skill. For that, you have to look at whether the world is sufficiently regular to support skill, which is true for chess masters and for recognizing the emotion in your wife's voice but probably isn't true in the stock market. Second, you have to ask whether the individual has had sufficient practice to acquire this skill. So confidence is not it. You've got to look from the outside. When a person makes a judgment, you have to ask what are the probabilities that this judgment is well-founded given the nature of the world in which that individual operates and the nature of the practice that the individual has had.

Ed Baker: Interesting, but there certainly are contexts in which confidence plays a dominant role in success, for example, in a leadership setting.

Daniel Kahneman: Absolutely. We reward confident optimists. There is no question that, in the context of leadership, somebody with high confidence is more likely to inspire trust in others and is more likely to attract resources that are needed for success. Optimism also facilitates resiliency in the context of execution. However, we need to distinguish situations in which optimism and confidence are useful from situations in which they are not. Roughly speaking, confidence is very useful in the context of execution, that is, when you are already committed to a course of action, you need to believe that you can do it. That will make you more resilient if things go badly, and thereby improve the real chances of success. If I have a favorite football team, I would like those players to be optimistic when they are on the field. In the context of decision making, however, I have absolutely no interest in my financial advisor being an optimist. I would like him to be as well-calibrated as possible.

Mark Anson: I've had experience working with pension funds over the years, and it's interesting to observe the group psychology and herding[7] that you see associated with large institutional investors. At least I've observed it from time to time with pension funds tending to move in the same direction at the same time. I noticed in your book that you talk about System 1 versus System 2 and the behavioral biases that can impact either of them. I was curious, from your point of view, do you find more behavioral bias embedded in a System 1 process versus a System 2? It seems like a System 2 process, which you refer to as a bit more analytical, might at times have the potential to be lazy and just accept what the rest of the herd is doing. Can you comment on that?

Daniel Kahneman: The way I analyze this in the book, most actions involve both systems. That is, System 1 quite often is the one that originates an idea or an impulse for an action. Then System 2 quite often endorses it, without checking sufficiently. That happens a great deal. In addition, System 2 quite often lacks the necessary knowledge. So you can slow yourself down, but mobilizing System 2 won't do anything for you if you don't have the tools to understand the situation. Slowing down is good when it allows you to deal with a situation more intelligently. Slowing down won't help when you are out of your depth.

Mark Anson: When people slow down, doesn't that tend to mean that they fall back in with the pack again, in that herding behavior that many have written about?

Daniel Kahneman: I'm not at all sure of that. I would attribute herding to a System 1 tendency. In situations of very high ambiguity, and when you have lost your confidence in your own ability to understand the world, then the tendency to do just what other people are doing is extremely powerful. It's also reinforced by social norms and by groups. If you see other pension funds doing something and you don't do it, you will get severely punished if you lose for not following the herd. So following the herd has an element of safety in it, and it's bound with System 1. I don't think of it as primarily a System 2 process. Herding is not necessarily something one does as the result of analysis. It is what one does when one's confidence is impaired.

Meir Statman: There are two areas that I hope you will not end this interview without addressing. One has to do with your work on well-being, and the other has to do with your work on fairness. Why do people with billions of dollars—hedge fund managers as one example—want even more money? I know what it does to their wealth, but what does it do to their well-being? Is it possible that a good part of what financial advisors do is increase investors' well-being while potentially diminishing their wealth?

Daniel Kahneman: Those are two very different questions, so I'll take them one at a time. We know from recent research that, beyond a certain income threshold, which is actually quite low—it's about $70,000 per household, emotional happiness doesn't seem to increase at all. Now, life satisfaction probably increases reasonably steadily with wealth. When people seek more wealth, although they will never spend what they already have, this is clearly because money is a proxy for something else. I mean, money is a proxy for ego satisfaction. So most of these people are in it because they need success, and money is just an index of success. That, I think, is the motivation for many people. Actually, I think the people who are strictly motivated by money rather than by success are mainly the poor and the very poor. For most of us professionals, money is a proxy for something else, and that is certainly true for hedge fund managers. So that's an answer to your first question.

Your second question is a very interesting one—what is the relationship between financial advising and the client's well-being? Actually, I've worked with that question before. In fact, with a well-respected investment advisory firm, Andrew Rosenfield[8] and I were involved in devising a program for advising very wealthy investors. There you're really more concerned with their well-being than with their wealth. Primarily you want to protect them from regret, you want to protect them from the emotions associated with very big losses. So you end up focusing more on their emotions than on their wealth.

Meir Statman: Can you give an example of how you might have done this?

Daniel Kahneman: That relates to another question, that is, how does one identify risk tolerance? Our thinking on this was that the issue is not so much tolerance for risk as it is tolerance for losses. Tolerance for losses means that you have to know—the individual investor has to know and certainly the advisor has to try to know—how much loss the person will be able to tolerate before he changes his mind about what he is doing. Clearly, changing course by and large is not a good idea, and selling low and buying high is not a good idea. You have to anticipate regret and identify the individuals who are very prone to regret. They're not going to be very good clients for the financial advisor. If people are very prone to regret, then you have to help them devise a plan that will minimize their regret. For the very wealthy, emotion is clearly important in determining what policy is appropriate.

Geoff Gerber: I remember hearing Amos Tversky present the findings of your collaborative research at a University of California, Berkeley, seminar on finance back in the early 1980s. He introduced the concept of loss aversion bias[9] that you're talking about, which, as you say, is the tendency to fear losses more than we value gains. The question from an investment manager's perspective or an investor's perspective is does the implementation of stop-loss limits[10] help alleviate the loss aversion bias?

Daniel Kahneman: The main question that I have found useful to ask when someone is very wealthy is how much loss is the individual willing to tolerate? That is, what fraction of their wealth are they actually willing to lose? It turns out that fraction is usually not very large. That's a very important parameter. How much do they really want to protect as much as possible, and how much are they willing to consider losing? That varies a lot among individuals. By and large, the very wealthy mostly want to protect their wealth, and they're willing to play with a small fraction of it. That is the fraction they are prepared to lose, but it's not a large fraction. So they're loss averse, not risk averse as such.

Geoff Gerber: So you're suggesting that setting a stop-loss higher or lower depends on your willingness to accept a loss? Is that a good approach?

Daniel Kahneman: For the individual who is very concerned about losses, I think this is certainly a good approach. That's the major question you want to ask the investor. How much are you willing to lose? Then you have to take steps so that they won't lose more than they are willing to lose. That's in effect stop-loss policy.

Ed Baker: Could you in fact organize questions that involve costs of insurance to see how much they'd be willing to pay for insurance that would protect against losses?

Daniel Kahneman: That's interesting. I hadn't thought of it that way—in terms of insurance. Yes, that would be an interesting approach. Also, people have to become aware of the fact that by stopping their losses, they are giving up some potential upside. Looking at the trade-off between the upside and the downside gives you a sense of their attitude toward losses and what you should encourage them to do.

Meir Statman: You mentioned that people are willing to play with or lose some portion of their money. I don't know if you have in mind that they keep two mental accounts: one is money that is not to be lost, and the other is money that can possibly be lost?

Daniel Kahneman: That is exactly what we have in mind. We actually had the individual construct two portfolios. One is a portfolio that is designed mainly for safety, and the other portfolio is designed to take advantage of opportunities. The relative size of the two portfolios represents one way of identifying loss aversion because with your riskier portfolio, that's an amount you can consider losing. It's not only two mental accounts. At least with some clients, we make this completely explicit, that is, clients receive information about two accounts, about their safe account and their riskier account. This is a very natural way for people to think.

Meir Statman: If I might move on to the issue of fairness, where you've done a lot of work, perhaps I can frame my question in the context of the fees that are charged by advisors. I think that financial advisors have more difficulty than other professionals, say physicians, lawyers, or accountants, in setting fees and justifying their fees. Advisors seem to be forever trying to hide their fees in one form or another. Can you speak to this issue of fairness?

Daniel Kahneman: Actually, this is a topic I haven't thought about, so I don't have a clear sense. In part, the need to hide fees comes from the fact that many of the advisors are frequently conflicted to some extent because if they're associated with a firm that provides products, then there is a problem associated with fees. Advisors who are completely hands-off, that is, those who are not involved with the products they are selling, probably should have no difficulties explaining their fees and charging for their services. It's those who are in a more ambiguous position who are probably sensitive about their fees. I haven't seen much discussion of the fairness of fees because clearly this is a competitive market, and there is enough variability in the fees for individuals to make their own choices.

Ed Baker: Moving on to a different area, I was interested in asking about your new work in adversarial collaboration.[11] I found that to be a fascinating turn of events in your life. What motivated that? Have you found some interesting opportunities to do new creative work? How can this be applied? It seems that if you could develop some systematic rules, it could be a major breakthrough in the way negotiations work. I'm thinking, of course, in the area of government.

> **"I was just very struck by how totally wasteful this is, because in all these exchanges nobody admits to having made an error. It is very striking, and quite frequently it becomes an exercise in sarcasm."**

Daniel Kahneman: I got into adversarial collaboration because there is a system in the scholarly literature where people critique other people's writings, and then there is a reply, and then there is a rejoinder. That's the routine in scientific publications. I was just very struck by how totally wasteful this is, because in all these exchanges nobody admits to having made an error. It is very striking, and quite frequently it becomes an exercise in sarcasm. It's just foul actually. So having been involved in some controversy, I became very interested in the possibility of trying to meet people who don't agree with me. All of us have a shared commitment to science, and we—at least in principle—also have a shared commitment to truth. That gives us some basis for working together to achieve truth. Now it turns out that even among scientists, the commitment to truth is—well, it's a real commitment—but emotion comes in. One of the striking things about adversarial collaboration—and I've had several—is that at the end of the collaboration, nobody feels that he has changed his mind much. That's very typical.

You asked whether adversarial collaboration could be implemented in politics. The question is whether there is enough of a shared commitment, a shared goal, for people to be interested in searching for compromise or in searching for joint action. This clearly exists among scientists, but it's much less likely to exist among true adversaries in the political domain, except possibly in a situation of crisis when it would become natural for adversaries to collaborate. I'm not very optimistic that adversarial collaboration can generally be extended to areas other than science. I've had luck with it. I've had good experiences with adversarial collaboration, I've avoided lifelong quarrels, and I have made friends. In sum, my experience has been a good one, but adversarial collaboration takes a lot of time and a lot of patience. It also sometimes takes quite a bit of self-control not to lose your temper with somebody who seems stuck on refusing to see the truth as you see it. So it's a mixed bag of experiences, and I'm not sure how far it can go beyond science.

Let me add that there are two practices that quite probably can advance or spread beyond science. One is, almost as a technique, to encourage adversaries to take each other's point of view and to make a speech that is, as it were, for the other side. That induces empathy, and it really helps you to understand what the other side is doing. That's a very worthwhile exercise if you're really interested in advancing cooperation. The other practice that seems really useful is socializing. I think one of the disasters in Washington is that apparently there is now very little socializing across political parties, whereas thirty or forty years ago, it was a rule that adversaries would drink and smoke together and go to football games together and so on. That is enormously important to mitigate adversarial relations, and we don't have that now.

Meir Statman: In politics, persuasion is the thing. It's less a matter of finding the truth than getting people to vote for you. I think there is an equivalent of that in the financial services industry, exploiting cognitive errors rather than countering them. For example, we see advertising that magnifies people's overconfidence in their ability to beat the market, rather than tamp it down. Can you speak to that?

Daniel Kahneman: Obviously, there is a lot of pandering to System 1 in advertising. I don't know if you have in mind the ads that encourage you to trade so as to beat the market and become rich. Those ads are clearly directed at overconfident people, and are intended to enhance their overconfidence. Most of advertisement is addressed to System 1, not to System 2. There is very little information in advertising, and anybody who watches programs with loads of advertising, such as the Super Bowl for example, would be hard put to find any information about any product. It is very striking—there is none. It's all appealing to different types of emotions.

Meir Statman: By one reliable estimate, U.S. investors would save more than $100 billion each year if they switched to low-cost index funds. Why aren't more investors using index funds? Why aren't they more sensitive to the fees involved?

"There is very little information in advertising, and anybody who watches programs with loads of advertising, such as the Super Bowl for example, would be hard put to find any information about any product. It is very striking—there is none. It's all appealing to different types of emotions."

Daniel Kahneman: I think that most people believe they are in the market to beat the market. If they are planning to beat the market, they are willing to pay some price. If, in your imagination, you're going to beat the market by a lot, then you become insensitive to fees. In order to become sensitive to fees, it's almost a precondition to accept that you're very unlikely to beat the market systematically, and that's a difficult realization for many investors. That relates to the other question of why aren't all investors in index funds. Clearly, there has been an increase in the amount of money invested in index funds, but I read the statistic of 25 percent of assets somewhere. Is that correct?

Meir Statman: At most, I would say.

Daniel Kahneman: This is clearly overconfidence at work, and to some extent the people who are selling these services are themselves overconfident. I had a marvelous experience many years ago with a financial advisor, whom I actually left—well, I had already left him when we had this conversation. I had moved to a safer line of investments, and he called me and said: "Look, what you are doing is stupid. We could make a lot of money for you. You are limiting your gains to a fixed amount, and last year we had several funds that did so much better than that amount." Then I looked back at the letter he had written me a year earlier in which he recommended specific funds. None of the funds he had recommended was among those that actually made a lot of money a year later. But he didn't know it. He had no interest in lying to me, because I had already left him and he knew I wasn't coming back. He was fighting for his own overconfidence. I think there's much more sincere overconfidence than lying among the professionals who think they can beat the market, and so they convince investors, and investors think, "Here is a guy with a track record of five winning years," and off they go.

Meir Statman: Obviously, cognitive errors get in the way, because the financial services industry is a great puzzle. In a world where people are smart—even if not rational—all would move on to index funds. The question I come back to is the question of well-being. Is it possible that we underestimate the joy that people derive from attempts to beat the market? Or that we underestimate the desire for the hope of getting rich through their investments?

Daniel Kahneman: I see the question you are raising, and it's a very interesting one. Clearly when people go to Las Vegas to gamble, most of them are not going to get rich, and they know that they are more likely to lose than to win, but they are going for the entertainment and the excitement and the thrill and the possibility of winning. Whether people who are investing think of it as going to Las Vegas, I personally doubt it very much. I don't think it's the same thing. They don't know that they're gambling—they think they're playing a game of skill.

Ed Baker: However, there are examples such as Warren Buffett,[12] and people see someone like that apparently making money consistently. Do they just misassess the probability of winning? Is that really what's going on?

Daniel Kahneman: I think so. Clearly from the examples you see or read about, there are successful people. If you went by the proportion of successful and unsuccessful people that you see in the media or that you hear talked about, then success overwhelms failure. Anybody who relies on what we call the availability heuristic[13] is going to find support for his overconfidence. That's overconfidence, not a search for well-being. The few who are in the market for the sheer excitement of it probably gamble small amounts, and know that they are in Las Vegas.

Meir Statman: Perhaps, but if you ask people who drive a Lexus or Rolls Royce if they do it for status, they would surely deny it. They would say it's because of the car's high quality and so on. I wonder if investors lack introspection about their wants.

Daniel Kahneman: To some extent, I think you are right. There are two separate questions. First, do people know the odds? The best evidence suggests that they don't know the odds, but they are truly optimistic about the likelihood of their winning. Second, when they play, when they are in the market, do they by and large derive well-being from it? Well, that's a complicated question, because if somebody is more sensitive to losses than to gains, then they don't get much well-being from the winning and losing. They get some excitement, and they quite possibly are deluded about how much they are winning and losing. That is, people have selective memories for their successes and failures, and they may actually misremember their previous record and think that it is better than it really was.

Margaret Towle: It's similar to 2008. When you ask people, nobody lost any money then.

Meir Statman: I lost money, I can assure you.

Margaret Towle: We've covered a wide range of topics so far today, Dr. Kahneman, but are there other areas of interest that you think are especially relevant when it comes to the investment industry as far as potential areas of research or areas that are unexplored now, given your conceptual framework?

Daniel Kahneman: Of course, there are many questions about the future, the future of research, and so on. I don't believe in long-run forecasting, and I don't believe that you can say the field is going in one direction or another. I have very little to say about where the field is going. Short-term, you can tell there is going to be more neuro-economics—that's fairly clear, because so many bright students are going into that field. The role of emotion in decision making is going to be discussed, and there's going to be more of it in the near future. Long-term, who knows?

Meir Statman: One sentence in *Thinking, Fast and Slow* that struck and delighted me was one where you said that you cringe when you hear people say that Amos Tversky and you proved that people are irrational. Could you elaborate on that? What is your sense of rationality? What does irrationality mean to you? I know that I have been using the term "normal" to define the opposite of rational.

Daniel Kahneman: I'm delighted with that question, and I'm actually very pleased to talk about that. The word rational[14] for me is a technical term. Rationality is defined in decision theory[15] as logical coherence, and it's very easy to test. In fact, a significant amount of research—and the research done by Amos and me, specifically—was dedicated to showing that people are not rational by that definition. But to call people irrational makes me cringe because the meaning of irrationality is associated for most people with emotion, with impulsivity, with frothing at the mouth. Our research was concerned with cognitive biases; we did not deal with mistakes that people make that arise from emotional impulsivity. As I understand the word, what we studied was not irrationality. I see a lot of System 1 influence, and System 1 is the emotional one, but I don't see all that much irrationality.

Ed Baker: On the other hand, you've resisted defining rationality, you said. If you were forced to come up with a definition, what would it be?

Daniel Kahneman: I think I just defined it. I accept the definition of rationality as a technical term. I don't use the word rationality except as that technical term. I don't say people are irrational. I speak of reasonableness, I follow Richard Thaler in talking about Econs[16] versus Humans, and I think Meir's use of normal is the same general idea. I just don't use the word much, except in its technical meaning. The so-called rational agent[17] hypothesis is outlandish and completely implausible. No finite mind could satisfy the requirement of rationality. The bottom line is that I don't need to define rationality, because it's defined as a technical term.

Ed Baker: Is there some underlying condition, though, that leads to efficient markets?

Daniel Kahneman: I don't know enough economics to answer that question. I could quote second-hand or third-hand that it doesn't take many rational agents to have enough money to enforce market discipline and so on. But I don't really know enough.

Meir Statman: Can you elaborate on what you said in your book about prospect theory?[18] You noted what prospect theory did to counter expected utility theory,[19] but you also pointed out the shortcomings of prospect theory in being true to reality. I'm not sure if I'm quoting it correctly, but I have this quote in my mind from Amos Tversky that "elegance is for tailors."[20]

Daniel Kahneman: Amos attributed that quote to Albert Einstein. I don't know if he was right—I never checked.

Meir Statman: Any quote where we don't know the source, we attribute to either [John Maynard] Keynes[21] or Einstein. In any event, would you comment on the fascination we have with higher mathematics and formal models and the field's direction in terms of how it expresses itself? I know you don't forecast long-term, but perhaps short-term?

Daniel Kahneman: Clearly, people who know mathematics have an advantage over people who don't, because they speak a language that others don't understand, whereas psychologists, sociologists, and people in professions such as that—most of the social scientists—speak in a language that, even if they use a little jargon, everybody can understand. So mathematics is an exclusive club, and there is a certain pride in belonging to it. It creates a mystique, and those who belong probably get a little more respect than they deserve. On the other hand, I have seen examples where clear mathematical

> "I was just very struck by how totally wasteful this is, because in all these exchanges nobody admits to having made an error. It is very striking, and quite frequently it becomes an exercise in sarcasm."

thinking really improves the quality of psychological theory. Amos Tversky was a master at it. He could use mathematics to think better. That's not true of all mathematical psychologists, but Amos really used mathematics to make himself think more clearly. There are other examples as well. In behavioral finance, for example, we have the demonstrations by Nicholas Barberis[22] of Yale University that one needs not only loss aversion but also narrow framing[23] in order to explain the behavior of individuals in the market. That was mathematical reasoning. It can be very fruitful when used in conjunction with good psychological intuition, so it is a very powerful tool.

Margaret Towle: We're nearing the end of our time, Dr. Kahneman, so I'll ask you if there's anything we haven't covered that you'd like to discuss.

Daniel Kahneman: No, we have covered more than I know.

Margaret Towle: Well, that's due to collective intelligence, I think, as far as the great questions that the group asked.

Ed Baker: I have one final question as to whether you have any thoughts for the regulators. Could any of your more recent work in behavioral finance, in particular, be helpful in forming market regulation?

Daniel Kahneman: I think there is no question about that. There are direct implications of behavioral economics and of the idea of bounded rationality[24] for regulation. The idea of the rational agent model has two pernicious consequences. One is that you don't need to protect consumers from themselves because they are rational, and therefore can be trusted to make the choices that are best for them. So you can oppose Social Security on the dual assumptions that people are rational and that they should bear the consequences of their actions. However, I believe that regulation is essential to protect people from predictable mistakes. You have to do that without abridging freedom, of course, but that can be done. And then you need to protect consumers from actors in the market that would deliberately exploit people's ignorance and their intellectual sloth.

Margaret Towle: This has been a most interesting discussion. We really appreciate your taking the time to share your views and talk with us. Thank you, Dr. Kahneman.

Daniel Kahneman: Thank you.

Endnotes

1 Richard H. Thaler (1945–) is an economist and professor of behavioral science and economics at The University of Chicago Booth School of Business. He is best known as a pioneering theorist in behavioral finance and for his collaboration with Daniel Kahneman and others in further defining the field of behavioral economics and finance.

2 Amos Tversky (1937–1996) was a cognitive and mathematical psychologist, a key figure in the discovery of systematic human cognitive bias and handling of risk, and a longtime collaborator of Daniel Kahneman. Their early work together focused on the psychology of prediction and probability judgment. The two went on to develop prospect theory, which endeavors to explain irrational human economic choices and is considered one of the seminal works of behavioral economics. Six years after Tversky's death, Dr. Kahneman received the 2002 Nobel Memorial Prize in Economic Sciences for the work he did in collaboration with Tversky. (The prize is not awarded posthumously.) Kahneman told the *New York Times* in an interview soon after receiving the honor (November 5, 2002): "I feel it is a joint prize. We were twinned for more than a decade."

3 In psychology, dual process theory is used to explain how a phenomenon can occur in two different ways or as a result of two different processes (and in various mixtures of the two): an implicit (or automatic) unconscious process and an explicit (or controlled) conscious process. Daniel Kahneman further differentiated these two styles of processing as System 1 and System 2. System 1 (or intuition) is rapid, automatic, and effortless, usually with strong emotional bonds included in the reasoning process. System 2 (or reasoning) is slower, deliberate, and subject to conscious judgments and attitudes.

4 The law of small numbers describes the judgmental bias that can occur when an assumption is made that the characteristics of a sample population can be estimated from a small number of observations or data points.

5 Heuristics of judgment are principles or methods used to potentially simplify assessments or judgments of probability. In psychology, heuristics are simple, efficient rules, hard-coded by evolutionary processes or learned, that are used to explain how people make decisions and solve problems, usually when facing complex situations or incomplete information. Although these rules work well under most circumstances, they can lead to systematic errors or cognitive biases. Examples of heuristics of judgment include representativeness, availability, and anchoring.

6 Gary A. Klein (1944–) is a research psychologist noted for pioneering the field of naturalistic decision making, focusing on the ability of intuition to support human decision making in high-pressure circumstances, such as firefighting and medical emergencies.

7 Herding describes the phenomenon of individuals in a group unconsciously acting together without planned direction.

8 Andrew Rosenfield, an economist and attorney, is managing partner of Guggenheim Partners, a financial services firm that provides wealth and investment management services to high-net-worth clients, foundations, and endowments. He also is managing partner and chief executive officer of The Greatest Good.

9 In economics and decision theory, loss aversion bias is a form of cognitive bias that describes the tendency to strongly prefer avoiding losses to acquiring gains. Studies suggest that losses are twice as powerful psychologically as gains. Loss aversion was first convincingly demonstrated by Amos Tversky and Daniel Kahneman.

10 Stop-loss limits are orders placed with a broker to sell a security when it reaches a certain price. A stop-loss order is designed to limit an investor's loss on a security position in advance, minimizing emotional decision making.

11 Adversarial collaboration is described as "a good-faith effort by unlike minds to conduct joint research, critiquing each other in the service of an ideal of truth to which both can contribute" on Dr. Kahneman's TED speaker page, http://www.ted.com/speakers/daniel_kahneman.html.

12 Warren Buffett (1930–) is an American investor, philanthropist, and chairman and chief executive officer of Berkshire Hathaway. Often referred to as "the oracle of Omaha," he was ranked by *Forbes* as the third-wealthiest person in the world, with a net worth of $44 billion, in March 2012.

13 The availability heuristic is a thought process that uses the ease with which examples come to mind, or knowledge that is readily available, to make judgments about the probability of events. This can result in a cognitive bias because the frequency with which examples come to mind does not accurately reflect their actual probability.

14 In psychology, the term rational is used to denote the use of conscious thought processes to solve problems.

15 Decision theory involves identifying the values, uncertainties, and other issues relevant to decision making. By outlining a set of alternatives and their potential consequences, decision theory can be used to help individuals make better-informed decisions.

16 Econs, a term coined by Richard Thaler, are the imaginary efficient individuals found only in economic theory who are able to weigh multiple options, forecast the consequences of each, and make logical choices, as opposed to actual humans, who are illogical, prone to generalize, biased in favor of the status quo, and more concerned with avoiding loss than making gains.

17 In economics and decision theory, a rational agent, which can include individuals, companies, or computer programs, has clear preferences, models uncertainty using expected values, and always chooses to perform the action that results in the optimal outcome for itself from among all feasible actions.

18 Prospect theory describes the ways in which individuals make choices among probabilistic alternatives that involve risk or uncertainty and evaluate potential losses and gains. Prospect theory, which attempts to model real-life choices rather than optimal decisions, holds that individuals make decisions based on the potential value of losses and gains (loss aversion) rather than the final outcome. The theory was developed by Daniel Kahneman and Amos Tversky in 1979 as a psychologically more accurate description of preferences versus expected utility theory.

19 In decision making, expected utility theory, which is based on elementary rules of rationality, addresses the analysis of choices among risky or uncertain prospects by measuring the value of various outcomes relative to respective probabilities, with the focus on the final outcome.

20 "If you are out to describe the truth, leave elegance to the tailor." Attributed to Albert Einstein (1879–1955) as well as to Ludwig Boltzmann (1844–1906), an Austrian physicist noted for advocating for atomic theory at a time when it was still controversial.

21 John Maynard Keynes (1883–1946) was a world-renowned British economist whose ideas, known as Keynesian economics, had a major impact on theories of modern economics and politics as well as on government fiscal policies.

22 Nicholas C. Barberis (1971–) is a professor of finance at the Yale School of Management, where his research focuses on behavioral finance, specifically using cognitive psychology to understand the pricing of financial assets.

23 Framing refers to the context in which a decision is made. An investor is said to use narrow framing when he makes an investment decision without considering the context of his total portfolio. Together, narrow framing and loss aversion may provide a method for understanding how individuals evaluate stock market risk by examining their evaluation of risk in experimental settings.

24 In decision making, bounded rationality holds that the rationality of individuals is limited by the information they possess, their cognitive limitations, and the finite amount of time available to make a decision. Economic models typically assume that the average person is rational and will, in large enough numbers, act according to preferences. The concept of bounded rationality revises this assumption to account for the fact that perfectly rational decisions are, in practice, often unfeasible because of the finite computational resources available for making them. Daniel Kahneman has proposed bounded rationality as a model to overcome some of the limitations of the rational agent model in economic literature.

References

Kahneman, Daniel. 2002. Autobiography. http://www.nobelprize.org/nobel_prizes/economics/laureates/2002/kahneman-autobio.html.

———. 2003. Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review* 93, no. 5 (December): 1,449–1,475.

———. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, Daniel, and Angus Deaton. 2010. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Science*s 107, no. 38 (September 21): 16,489–16,493.

Kahneman, Daniel, and Edwin E. Ghiselli. 1962. Validity and Nonlinear Heteroscedastic Models. *Personnel Psycholog*y 15, no. 1 (March): 1–11.

Kahneman, Daniel, and Gary Klein. 2009. Conditions for Intuitive Expertise: A Failure to Disagree. *American Psychologist* 64, no. 6 (September): 515–526.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometric*a 47, no. 2 (March): 263–291.

Thaler, Richard H., Amos Tversky, Daniel Kahneman, and Alan Schwartz. 1997. The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test. *Quarterly Journal of Economics* 112, no. 2 (May): 647–661.

Tversky, Amos, and Daniel Kahneman. 1971. Belief in the Law of Small Numbers. *Psychological Bulletin* 76, no. 2: 105–110.

———. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, no. 4157 (September 27): 1,124–1,131.

Journal of Economic Literature 2012, 50(4), 1–12
http://dx.doi.org/10.1257/jel.50.4.1

Psychologists at the Gate: A Review of Daniel Kahneman's *Thinking, Fast and Slow*

Andrei Shleifer*

The publication of Daniel Kahneman's book, Thinking, Fast and Slow, *is a major intellectual event. The book summarizes, but also integrates, the research that Kahneman has done over the past forty years, beginning with his path-breaking work with the late Amos Tversky. The broad theme of this research is that human beings are intuitive thinkers and that human intuition is imperfect, with the result that judgments and choices often deviate substantially from the predictions of normative statistical and economic models. In this review, I discuss some broad ideas and themes of the book, describe some economic applications, and suggest future directions for research that the book points to, especially in decision theory.* (*JEL* A12, D03, D80, D87)

1. *Introduction*

The publication of Daniel Kahneman's book, *Thinking, Fast and Slow* (Farrar, Straus, and Giroux 2011), is a major intellectual event. The book summarizes, but also integrates, the research that Kahneman has done over the past forty years, beginning with his path-breaking work with the late Amos Tversky. The broad theme of this research is that human beings are intuitive thinkers and that human intuition is imperfect, with the result that judgments and choices often deviate substantially from the predictions of normative statistical and economic models. This research has had a major impact on psychology, but also on such diverse areas of economics as public finance, labor economics, development, and finance. The broad field of behavioral economics—perhaps the most important conceptual innovation in economics over the last thirty years—might not have existed without Kahneman and Tversky's fundamental work. It certainly could not have existed in anything like its current form. The publication of Kahneman's book will bring some of the most innovative and fundamental ideas of twentieth century social science to an even broader audience of economists.

In this review, I discuss some broad ideas and themes of the book. Although it would be relatively easy to carry on in the spirit of

* Department of Economics, Harvard University. I have benefited from generous comments of Nicholas Barberis, Pedro Bordalo, Thomas Cunningham, Nicola Gennaioli, Matthew Gentzkow, Owen Lamont, Sendhil Mullainathan, Josh Schwartzstein, Jesse Shapiro, Tomasz Strzalecki, Dmitry Taubinsky, Richard Thaler, and Robert Vishny. They are not, however, responsible for the views expressed in this review. I do not cite specific papers of Kahneman when the material is described in the book.

the first paragraph, constrained only by my limited vocabulary of adjectives, I will seek to accomplish a bit more. First, because the book mentions few economic applications, I will describe some of the economic research that has been substantially influenced by this work. My feeling is that the most profound influence of Kahneman and Tversky's work on economics has been in finance, on what has now become the field of behavioral finance taught in dozens of undergraduate and graduate economics programs, as well as at business schools. I learned about Kahneman and Tversky's work in the 1980s as a graduate student, and it influenced my own work in behavioral finance enormously.

Second, I believe that while Kahneman and Tversky's work has opened many doors for economic research, some of the fundamental issues it has raised remain work in progress. I will thus discuss what Kahneman's work suggests for decision theory, primarily as I see it through the lens of my recent work with Nicola Gennaioli and Pedro Bordalo (Gennaioli and Shleifer 2010; Bordalo, Gennaioli, and Shleifer 2012a, 2012b, 2012c).

Before turning to the book, let me briefly address the two common objections to the introduction of psychology into economics, which have been bandied around for as long as the field has existed. The first objection holds that, while psychological quirks may influence individual decisions at the boundary, the standard economic model describes first order aspects of human behavior adequately, and economists should focus on "first order things" rather than quirks. Contrary to this objection, DellaVigna (2009) summarizes a great deal of evidence of large and costly errors people make in important choices. Let me illustrate. First, individuals pay large multiples of actuarially fair value to buy insurance against small losses, as well as to reduce their deductibles (Sydnor 2010). In the standard model, such choices imply astronomical levels of risk aversion. Second, the standard economic view that persuasion is conveyance of information seems to run into a rather basic problem that advertising is typically emotional, associative, and misleading—yet nonetheless effective (Bertrand et al. 2010; DellaVigna and Gentzkow 2010; Mullainathan, Schwartzstein, and Shleifer 2008). Third, after half a century of teaching by financial economists that investors should pick low-cost index funds, only a minority do, while most select high-cost actively managed funds that underperform those index funds. These kinds of behavior matter for both prices and resource allocation. Explaining such behavior with the standard model is possible, but requires intellectual contortions that are definitely not "first order."

The second objection holds that market forces eliminate the influence of psychological factors on prices and allocations. One version of this argument, made forcefully by Friedman (1953) in the context of financial markets, holds that arbitrage brings prices, and therefore resource allocation, to efficient levels. Subsequent research has shown, however, that Friedman's argument—while elegant—is theoretically (and practically) incorrect. Real-world arbitrage is costly and risky, and hence limited (see, e.g., Grossman and Miller 1988, DeLong et al. 1990, Shleifer and Vishny 1997). Dozens of empirical studies confirm that, even in markets with relatively inexpensive arbitrage, identical, or nearly identical, securities trade at different prices. With costlier arbitrage, pricing is even less efficient.

A second version of the "forces of rationality" objection holds that participants in real markets are specialists invulnerable to psychological quirks. List's (2003) finding that professional baseball card traders do not exhibit the so-called endowment effect is supportive of this objection. The problem with taking this too far is that individuals make lots of critical decisions—how much to save, how to invest, what to buy—on their own, without experts. Even when people receive expert help, the incentives of experts are often to take advantage of psychological biases of their customers. Financial advisors direct savers to expensive, and often inappropriate, products, rather than telling them to invest in index funds (Chalmers and Reuter 2012; Gennaioli, Shleifer, and Vishny 2012). Market forces often work to strengthen, rather than to eliminate, the influence of psychology.

2. *System 1 and System 2*

Kahneman's book is organized around the metaphor of System 1 and System 2, adopted from Stanovich and West (2000). As the title of the book suggests, System 1 corresponds to thinking fast, and System 2 to thinking slow. Kahneman describes System 1 in many evocative ways: it is intuitive, automatic, unconscious, and effortless; it answers questions quickly through associations and resemblances; it is nonstatistical, gullible, and heuristic. System 2 in contrast is what economists think of as thinking: it is conscious, slow, controlled, deliberate, effortful, statistical, suspicious, and lazy (costly to use). Much of Kahneman and Tversky's research deals with System 1 and its consequences for decisions people make. For Kahneman, System 1 describes "normal" decision making. System 2, like the U.S. Supreme Court, checks in only on occasion.

Kahneman does not suggest that people are incapable of System 2 thought and always follow their intuition. System 2 engages when circumstances require. Rather, many of our actual choices in life, including some important and consequential ones, are System 1 choices, and therefore are subject to substantial deviations from the predictions of the standard economic model. System 1 leads to brilliant inspirations, but also to systematic errors.

To illustrate, consider one of Kahneman and Tversky's most compelling questions/experiments:

> An individual has been described by a neighbor as follows: "Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." Is Steve more likely to be a librarian or a farmer?

Most people reply quickly that Steve is more likely to be a librarian than a farmer. This is surely because Steve resembles a librarian more than a farmer, and associative memory quickly creates a picture of Steve in our minds that is very librarian-like. What we do not think of in answering the question is that there are five times as many farmers as librarians in the United States, and that the ratio of male farmers to male librarians is even higher (this certainly did not occur to me when I first read the question many years ago, and does not even occur to me now as I reread it, unless I force myself to remember). The base rates simply do not come to mind and thus prevent an accurate computation and answer, namely that Steve is more likely to be a farmer. System 2 does not engage.

In another example (due to Shane Frederick), one group of respondents is asked (individually) to estimate the total number of murders in Detroit in a year. Another group is asked to estimate the total number of murders in Michigan in a year. Typically, the first group on average estimates a higher number of murders than the second. Again, System 1 thinking is in evidence. Detroit evokes a violent city, associated with many murders. Michigan evokes idyllic apple-growing farmland. Without System 2 engagement, the fact that Detroit is *in* Michigan does not come to mind for the second group of respondents, leading—across subjects—to a dramatic violation of basic logic.

Kahneman's other examples of System 1 thinking include adding 2 + 2, completing the words "bread and . . . ," and driving a car on an empty road. Calling all these examples System 1 thinking captures the rapid, intuitive, automatic response, which usually gets the right answer, but sometimes—as with Steve and murders in Michigan—does not. Yet unfortunately things are not as clear as they look, once we apply our own System 2 thinking to System 1.

First, as Kahneman readily recognizes, the domains of System 1 and System 2 differ across people. For most (all?) readers of this review, computing 20 × 20 is a System 1 effortless task, largely because economists have both been selected to be good at it and have had lots of practice. But for many people who are not experts, this operation is effortful, or even impossible, and is surely the domain of System 2. In contrast, screwing in a light bulb is very System 2 for me—conscious, effortful, and slow—but not so for most people, I gather. As people gain knowledge or expertise, the domains of the two systems change. In fact, the classification of decisions into products of System 1 and System 2 thinking seems to be even harder. Go back to murders in Detroit and in Michigan. The question surely evoked images of bombed-out Detroit and pastoral Michigan, but constructing the estimate also requires a substantial mental effort. Both systems seem to be in action.

Second, the challenge of going beyond the labels is that System 2 is not perfect, either. Many people would get 20 × 20 wrong, even if they think hard about it. The idea that conscious thought and computation are imperfect goes back at least to Herbert Simon and his concept of bounded rationality. Bounded rationality is clearly important for many problems (and in fact has been fruitfully explored by economists), but it is very different from Kahneman's System 1. Kahneman's brilliant insight—illustrated again and again throughout the book—is that people do not just get hard problems wrong, as bounded rationality would predict; they get utterly trivial problems wrong because they don't think about them in the right way. This is a very different notion than bounded rationality. Still, the challenge remains that when we see a decision error, it is not obvious whether to attribute it to System 1 thinking, System 2 failure, or a combination.

Third, the classification of thought into System 1 and System 2 raises tricky questions of the relationship between the two. Because System 1 includes unconscious attention, perception, and associative memory, much of the informational input that System 2 receives comes via System 1. Whether and how System 1 sends "up" the message if at all is a bit unclear. In other words, what prompts the engagement of System 2? What would actually trigger thinking about relative numbers of male librarians and farmers in the United States, or even whether Michigan includes Detroit? I am not sure that anything but a hint would normally do it. Perhaps System 2 is almost always at rest. Furthermore, one function of System 2 appears to be to "check the answers" of System 1, but if information "sent up" is incomplete and distorted, how would System 2 know? To strain the legal analogy a bit further, appellate courts in the United States must accept fact finding of trial courts as given, so many errors—as well as deliberate distortions—creep in precisely at the fact-finding trial stage, rather than in the appealable application of law to the facts. Kahneman writes that "the division of labor between System 1 and System 2 is highly efficient: it minimizes effort and optimizes performance" (25). I am not sure why he says so. If System 1 guides our insurance and investment choices described in the introduction, then System 2 seems rather disengaged even when the costs of disengagement are high.

To put these comments differently, each of System 1 and System 2 appears to be a collection of distinct mental processes. System 1 includes unconscious attention, perception, emotion, memory, automatic causal narratives, etc. I am worried that, once the biology of thought is worked out, what actually happens in our heads is unlikely to neatly map into fast and slow thinking. The classification is an incredibly insightful and helpful metaphor, but it is not a biological construct or an economic model. Turning metaphors into models remains a critical challenge.

3. *Heuristics and Biases*

One of the two main bodies of Kahneman and Tversky's work has come to be known as "Heuristics and Biases." This research deals, broadly, with intuitive statistical prediction. The research finds that individuals use heuristics or rules of thumb to solve statistical problems, which often leads to biased estimates and predictions. Kahneman and Tversky have identified a range of now famous heuristics, which fall into two broad categories.

Some heuristics involve respondents answering questions for which they do not have much idea about the correct answer, and must retrieve a guess from their memory. The problem given to them is not self-contained. As a consequence, respondents grasp at straws, and allow their answers to be influenced by objectively irrelevant frames. One example of this is the anchoring heuristic. A wheel of fortune, marked from 0 to 100, is rigged by experimenters to stop only at either 10 or 65. After a spin, students write down the number at which it stopped, and are then asked two questions: Is the percentage of African nations among U.N. members larger or smaller than the number you just wrote? What is your best guess of the percentage of African nations in the United Nations? For students who saw the wheel of fortune stop at 10, the average guess was 25 percent. For those who saw it stop at 65, the average guess was 45 percent. Similar experiments have been run with lengths of rivers, heights of mountains, and so on. The first question anchors the answer to the second. Kahneman interprets anchoring as an extreme example of System 1 thinking: planting a number in one's head renders it relevant to fast decisions.

The second category of heuristics is much closer to economics and, in fact, has received a good deal of attention from economists. These heuristics describe statistical problems in which respondents receive all the information they need, but nonetheless do not use it correctly. Not all available information seems to come to the top of the mind, leading to errors. Examples of neglected decision-relevant information include base rates (even when they are explicitly stated), low probability but nonsalient events, and chance. The finding that the causal and associative System 1 does not come up with chance as an explanation seems particularly important. Kahneman recalls a magnificent story of Israeli Air Force officers explaining to him that being tough with pilots worked miracles because, when pilots had a poor landing and got yelled at, their next landing was better, but when they had a great landing and got praised, their next landing was worse. To these officers, the role of chance and consequent mean reversion in landing quality did not come to mind as an explanation.

The best known problems along these lines describe the representativeness heuristic, of which the most tantalizing is Linda, here slightly abbreviated:

> Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

After seeing the description, the respondents are asked to rank in order of likelihood various scenarios: Linda is (1) an elementary school teacher, (2) active in the feminist movement, (3) a bank teller, (4) an insurance salesperson, or (5) a bank teller also active in the feminist movement. The remarkable finding is that (now generations of) respondents deem scenario (5) more likely than scenario (3), even though (5) is a special case of (3). The finding thus violates the most basic laws of probability theory. Not only do many students get the Linda problem wrong, but some object, sometimes passionately, after the correct answer is explained.

What's going on here? The description of Linda brings to mind, presumably from associative memory, a picture that does not look like a bank teller. Asked to judge the likelihood of scenarios, respondents automatically match that picture to each of these scenarios, and judge (5) to be more similar to Linda than (3). System 1 rather easily tells a story for scenario (5), in which Linda is true to her beliefs by being active in the feminist movement, yet must work as a bank teller to pay the rent. Telling such a story for (3) that puts all the facts together is more strenuous because a stereotypical bank teller is not a college radical. The greater similarity of Linda to the feminist bank teller leads respondents to see that as a more likely scenario than merely a bank teller.

Many studies have unsuccessfully tried to debunk Linda. It is certainly true that if you break Linda down for respondents (there are 100 Lindas, some are bank tellers, some are feminist bank tellers, which ones are there more of?)—if you engage their System 2—you can get the right answer. But this, of course, misses the point, namely that, left to our own devices, we do not engage in such breakdowns. System 2 is asleep. In Linda, as in Steve the librarian and many other experiments, the full statistical problem simply does not come to mind, and fast-thinking respondents—even when they do strain a bit—arrive at an incorrect answer.

There have been several attempts by economists to model such intuitive statistics (e.g., Mullainathan 2000, 2002; Rabin 2002; Rabin and Vayanos 2010; Schwartzstein 2012). In one effort that seeks to stay close to Kahneman's System 1 reasoning, Gennaioli and Shleifer (2010) argue that individuals solve decision problems by representing them—automatically but incompletely—in ways that focus on features that are statistically more associated with the object being assessed. In the Linda problem, the feminist bank teller is described comprehensively and hence represented as a feminist bank teller. A bank teller, in contrast, is not described comprehensively, and bank teller evokes the stereotype of a nonfeminist because not being a feminist is relatively more associated with being a bank teller than being a feminist. The decision-maker thus compares the likelihoods *not* of bank teller versus feminist bank teller, but rather of the stereotypical (representative) nonfeminist bank teller versus feminist bank teller, and concludes that Linda the college radical is more likely to be the latter. This approach turns out to account for a substantial number of heuristics discussed in Kahneman's book. The key idea, though, is very much in the spirit of System 1 thinking, but made tractable using economic modeling, namely that to make judgments we *represent* the problem automatically via the functioning of attention, perception, and memory, and our decisions are subsequently distorted by such representation.

The representativeness heuristic had a substantial impact on behavioral finance, largely because it provides a natural account of *extrapolation*—the expectation by investors that trends will continue. The direct evidence on investor expectations of stock returns points to a strong extrapolative component (e.g., Vissing-Jorgensen 2004). Extrapolation has been used to understand price bubbles (Kindleberger 1978), but also the well-documented overvaluation and subsequent reversal of high performing growth stocks (De Bondt and Thaler 1985; Lakonishok, Shleifer, and Vishny 1994). Indeed, data for a variety of securities across markets show that price trends continue over a period of several months (the so-called momentum), but that extreme performance reverts over longer periods (Cutler, Poterba, and Summers 1991). Even more dramatically, investors put money into well-performing mutual funds, into stock funds and stock market-linked insurance products after the stock market has done well (Frazzini and Lamont 2008; Yagan 2012). Such phenomena have been described colorfully as investors "jumping on the bandwagon" believing that "the trend is your friend," and failing to realize that "trees do not grow to the sky," that "what goes up must come down," etc.

Heuristics provide a natural way of thinking about these phenomena, and can be incorporated into formal models of financial markets (see, e.g., Barberis, Shleifer, and Vishny 1998). Specifically, when investors pour money into hot, well-performing assets, they may feel that these assets are similar to, or resemble, other assets that have kept going up. Many high tech stocks look like the next Google, or at least System 1 concludes that they do. Extrapolation is thus naturally related to representativeness, and supports the relevance of Kahneman's work not just in the lab, but also in the field.

4. *Prospect Theory*

Prospect Theory has been Kahneman and Tversky's most influential contribution, and deservedly so. In a single paper, the authors proposed an alternative to standard theory of choice under risk that was at the same time quite radical and tractable, used the theory to account for a large number of outstanding experimental puzzles, and designed and implemented a collection of new experiments used to elucidate and test the theory. In retrospect, it is difficult to believe just how much that paper had accomplished, how new it was, and how profound its impact has been on behavioral economics.

Prospect Theory rests on four fundamental assumptions. First, risky choices are evaluated in terms of their gains and losses relative to a reference point, which is usually the status quo wealth. Second, individuals are loss averse, meaning extremely risk averse with respect to small bets around the reference point. Third, individuals are risk averse in the domain of gains, and risk loving in the domain of losses. And finally, in assessing lotteries, individuals convert objective probabilities into decision weights that overweight low probability events and underweight high probability ones.

The first assumption is probably the most radical one. It holds that rather than integrating all risky choices into final wealth states, as standard theory requires, individuals frame and evaluate risky bets narrowly in terms of their gains and losses relative to a reference point. In their 1979 paper, Kahneman and Tversky did not dwell on what the reference point is, but for the sake of simplicity took it to be the current wealth. In a 1981 *Science* paper, however, they went much further in presenting a very psychological view of the reference point: "The reference outcome is usually a state to which one has adapted; it is sometimes set by social norms and expectations; it sometimes corresponds to a level of aspiration, which may or may not be realistic" (456). The reference point is thus left as a rather unspecified part of Kahneman and Tversky's theory, their measure of "context" in which decisions are made. Koszegi and Rabin (2006) suggest that reference points should be rational expectations of future consumption, a proposal that brings in calculated thought. Pope and Schweitzer (2011) find that goals serve as reference points in professional golf. Hart and Moore (2008) believe that contracts serve as reference points for future negotiations. A full elaboration of where reference points come from is still "under construction."

The second assumption of Prospect Theory is loss aversion. It is inspired by a basic and intuitively appealing experiment in which people refuse to take bets that give them a 60 percent probability of winning a dollar and a 40 percent probability of losing a dollar, even though such a refusal implies an implausibly high level of risk aversion (Rabin 2000). Kahneman justifies this assumption by noting that, biologically, losses might be processed in part in the amygdala in the same way as threats. Kahneman and Tversky modeled this assumption as a kink in the value function around the reference point. In fact, in its simplest version, Prospect Theory (without assumptions 3 and 4 described below) is occasionally presented graphically with a piecewise linear value function, with the slope of 1 above the origin and 2 below the origin (reference point), and a kink at the origin that captures loss aversion. Kahneman sees loss aversion as the most important contribution of Prospect Theory to behavioral economics, perhaps because it has been used to account for the endowment effect (the finding, both in the lab and in the field, that individuals have a much higher reservation price for an object they own than their willingness to pay for it when they do not own it).

The third assumption is that behavior is risk averse toward gains (as in standard theory) and risk seeking toward losses. It is motivated by experiments in which individuals choose a gamble with a 50 percent chance of losing $1,000 over a certainty of losing $500. This assumption receives some though not total support (Thaler and Johnson 1990), and has not been central to Prospect Theory's development.

The fourth assumption of Prospect Theory is quite important. That is the assumption of an inverted S-shaped function converting objective probabilities into decision weights, which blows up low probabilities and shrinks high ones (but not certainty). The evidence used to justify this assumption is the excessive weights people attach to highly unlikely but extreme events: they pay too much for lottery tickets, overpay for flight insurance at the airport, or fret about accidents at nuclear power plants. Kahneman and Tversky use probability weighting heavily in their paper, adding several functional form assumptions (subcertainty, subadditivity) to explain various forms of the Allais paradox. In the book, Kahneman does not talk about these extra assumptions, but without them Prospect Theory explains less.

To me, the stable probability weighting function is problematic. Take low probability events. Some of the time, as in the cases of plane crashes or jackpot winnings, people put excessive weight on them, a phenomenon incorporated into Prospect Theory that Kahneman connects to the availability heuristic. Other times, as when investors buy AAA-rated mortgage-backed securities, they neglect low probability events, a phenomenon sometimes described as black swans (Taleb 2007). Whether we are in the probability weighting function or the black swan world depends on the context: whether or not people recall and are focused on the low probability outcome.

More broadly, how people think about the problem influences probability weights and decisions. In one of Kahneman and Tversky's most famous examples, results from two potential treatments of a rare disease are described, alternatively, in terms of lives saved and lives lost. The actual outcomes—gains and losses of life—are identical in the two descriptions. Yet respondents choose the "safer" treatment when description is in terms of lives saved, and the "riskier" treatment when description is in terms of lives lost. The framing or representation of the problem thus changes probability weights even when objective outcomes are identical. In another study, Rottenstreich and Hsee (2001) show that decision weights depend on how "affect-rich" the outcomes are, and not just on their probabilities. Bordalo, Gennaioli, and Shleifer (2012c) present a model in which attention is drawn to salient, or unusual, payoffs. In their model, unlike in Prospect Theory, individuals overweigh only low probability events that are associated with extreme, or salient, payoffs. The model explains all the same findings as Prospect Theory, but also several additional ones, including preference reversals (people sometimes prefer *A* to *B*, but are willing to pay more for *B* than for *A* when considering the two in isolation). Kahneman of course recognizes the centrality of context in shaping mental representation of problems when he talks about the WYSIATI principle (what you see is all there is).

Prospect Theory is an enormously useful model of choice because it accounts for so much evidence and because it is so simple. Yet it achieves its simplicity by setting to one side both in its treatment of reference points and its model of probability weights precisely the System 1 mechanisms that shape how a problem is represented in our minds. For a more complete framework, we need better models of System 1.

Prospect Theory has been widely used in economics, and many of the applications are described in DellaVigna (2009) and Barberis (forthcoming). Finance is no exception. Benartzi and Thaler (1995) have argued, for example, that it can explain the well-known equity premium puzzle, the empirical observation that stocks on average earn substantially higher returns than bonds. Benartzi and Thaler observed that while stocks do extremely well in the long run, they can fall a lot in the short run. When investors have relatively short horizons and also, in line with Prospect Theory, are loss averse, this risk of short-term losses in stocks looms large, makes stocks unattractive, and therefore cheap, thus explaining the equity premium. More recently, Barberis and Huang (2008) argue that the probability weighting function of Prospect Theory has the further implication that investors are highly attracted to positive skewness in returns, since they place excessive weights on unlikely events. The evidence on overpricing of initial public offerings and out of the money options is consistent with this prediction.

5. *What's Ahead?*

In conclusion, let me briefly mention three directions in which I believe the ship launched by Kahneman and Tversky is headed, at least in economics. First, although I did not talk much about this in the review, Kahneman's book on several occasions discusses the implications of his work for policy. At the broadest level, how should economic policy deal with System 1 thinking? Should it respect individual preferences as distinct from those dictated by the standard model or even by the laws of statistics? Should it try to debias people to get them to make better decisions?

I have avoided these questions in part because they are extremely tricky, at both philosophical and practical levels (Bernheim and Rangel 2009). But one theme that emerges from Kahneman's book strikes me as important and utterly convincing. Faced with bad choices by consumers, such as smoking or undersaving, economists as System 2 thinkers tend to focus on education as a remedy. Show people statistics on deaths from lung cancer, or graphs of consumption drops after retirement, or data on returns on stocks versus bonds, and they will do better. As we have come to realize, such education usually fails. Kahneman's book explains why: System 2 might not really engage until System 1 processes the message. If the message is ignored by System 1, it might never get anywhere. The implication, clearly understood by political consultants and Madison Avenue advertisers, is that effective education and persuasion must connect with System 1. Calling the estate tax "the death tax" may work better to galvanize its opponents than statistics on hardworking American farmers who may have to pay. Thaler and Sunstein's (2008) *Nudge* advocates policies that simplify decisions for people relying on System 1 in situations, such as saving for retirement, where even an educated System 2 might struggle.

Beyond the changing thinking on economic policy, Kahneman's work will continue to exert a growing influence on our discipline. A critical reason for this is the rapidly improving quality of economic data from the field, from experiments, and from field experiments. Confronted with the realities of directly observed human behavior—financial choices made by investors, technology selection by farmers, insurance choices by the elderly—economists have come to psychology for explanations, especially to the work described in Kahneman's book. Rapidly expanding data on individual choices is the behavioral economist's best friend.

But it seems to me that some of the most important advances in the near future both need to come, and will come, in economic theory. Economics, perhaps like any other discipline, advances through changes in standard models: witness the enormous influence of Prospect Theory itself. In contrast, we do not have a standard model of heuristics and biases, and as I argued, Prospect Theory is still a work in progress. Fortunately, the broad ideas discussed in Kahneman's book, and in particular his emphasis on the centrality of System 1 thinking, provide some critical clues about the features of the models to come.

In particular, the main lesson I learned from the book is that we represent problems in our minds, quickly and automatically, before we solve them. Such representation is governed by System 1 thinking, including involuntary attention drawn to particular features of the environment, focus on these features, and recall from memory of data associated with these perceptions. Perhaps the fundamental feature of System 1 is that what our attention is drawn to, what we focus on, and what we recall is not always what is most necessary or needed for optimal decision making. Some critical information is ignored; other—less relevant—information receives undue attention because it stands out. In this respect, the difference from the models of bounded rationality, in which information is optimally perceived, stored, and retrieved, is critical. System 1 is automatic and reactive, not optimizing.

As a consequence, when we make a judgment or choice, we do that on the basis of incomplete and selected data assembled via a System 1-like mechanism. Even if the decisions are optimal at this point given what we have in mind, they might not be optimal given the information potentially available to us both from the outside world and from memory. By governing what we are thinking about, System 1 shapes what we conclude, even when we are thinking hard.

Kahneman's book, and his lifetime work with Tversky, had and will continue to have enormous impact on psychology, applied economics, and policy making. Theoretical work on Kahneman and Tversky's ideas has generally modeled particular heuristics and choices under risk separately, without seeking common elements. A potentially large benefit of Kahneman's book is to suggest a broader theme, namely that highly selective perception and memory shape what comes to mind before we make decisions and choices. Nearly all the phenomena the book talks about share this common thread. In this way, Kahneman points toward critical ingredients of a more general theory of intuitive thinking, still an elusive, but perhaps achievable, goal.

References

Barberis, Nicholas. Forthcoming. "Thirty Years of Prospect Theory in Economics." *Journal of Economic Perspectives*.

Barberis, Nicholas, and Ming Huang. 2008. "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices." *American Economic Review* 98 (5): 2066–2100.

Barberis, Nicholas, Andrei Shleifer, and Robert W. Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics* 49 (3): 307–43.

Benartzi, Shlomo, and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics* 110 (1): 73–92.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124 (1): 51–104.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman. 2010. "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment." *Quarterly Journal of Economics* 125 (1): 263–306.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012a. "Salience and Consumer Choice." National Bureau of Economic Research Working Paper 17947.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012b. "Salience in Experimental Tests of the Endowment Effect." *American Economic Review* 102 (3): 47–52.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012c. "Salience Theory of Choice Under Risk." *Quarterly Journal of Economics* 127 (3): 1243–85.

Chalmers, John, and Jonathan Reuter. 2012. What Is the Impact of Financial Advisors on Retirement Portfolio Choices and Outcomes?" National Bureau of Economic Research Working Paper 18158.

Cutler, David M., James M. Poterba, and Lawrence H. Summers. 1991. "Speculative Dynamics." *Review of Economic Studies* 58 (3): 529–46.

De Bondt, Werner F. M., and Richard H. Thaler. 1985. "Does the Stock Market Overreact?" *Journal of Finance* 40 (3): 793–805.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2): 315–72.

DellaVigna, Stefano, and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics* 2 (1): 643–69.

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy* 98 (4): 703–38.

Frazzini, Andrea, and Owen A. Lamont. 2008. "Dumb Money: Mutual Fund Flows and the Cross-Section of Stock Returns." *Journal of Financial Economics* 88 (2): 299–322.

Friedman, Milton. 1953. "The Case for Flexible Exchange Rates." In *Essays in Positive Economics*, 157–203. Chicago and London: University of Chicago Press.

Gennaioli, Nicola, and Andrei Shleifer. 2010. "What Comes to Mind." *Quarterly Journal of Economics* 125 (4): 1399–1433.

Gennaioli, Nicola, Andrei Shleifer, and Robert W. Vishny. 2012. "Money Doctors." National Bureau of Economic Research Working Paper 18174.

Grossman, Sanford J., and Merton H. Miller. 1988. "Liquidity and Market Structure." *Journal of Finance* 43 (3): 617–37.

Hart, Oliver, and John Moore. 2008. "Contracts as Reference Points." *Quarterly Journal of Economics* 123 (1): 1–48.

Kindleberger, Charles P. 1978. *Manias, Panics, and Crashes: A History of Financial Crises*. New York: Basic Books.

Koszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics* 121 (4): 1133–65.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny. 1994. "Contrarian Investment, Extrapolation, and Risk." *Journal of Finance* 49 (5): 1541–78.

List, John A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118 (1): 41–71.

Mullainathan, Sendhil. 2000. "Thinking through Categories." Unpublished.

Mullainathan, Sendhil. 2002. "A Memory-Based Model of Bounded Rationality." *Quarterly Journal of Economics* 117 (3): 735–74.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer. 2008. "Coarse Thinking and Persuasion." *Quarterly Journal of Economics* 123 (2): 577–619.

Pope, Devin G., and Maurice E. Schweitzer. 2011. "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes." *American Economic Review* 101 (1): 129–57.

Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica* 68 (5): 1281–92.

Rabin, Matthew. 2002. "Inference by Believers in the Law of Small Numbers." *Quarterly Journal of Economics* 117 (3): 775–816.

Rabin, Matthew, and Dimitri Vayanos. 2010. "The Gambler's and Hot-Hand Fallacies: Theory and Applications." *Review of Economic Studies* 77 (2): 730–78.

Rottenstreich, Yuval, and Christopher K. Hsee. 2001. "Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk." *Psychological Science* 12 (3): 185–90.

Schwartzstein, Joshua. 2012. "Selective Attention and Learning." http://www.dartmouth.edu/~jschwartzstein/papers/sa.pdf.

Shleifer, Andrei, and Robert W. Vishny. 1997. "The Limits of Arbitrage." *Journal of Finance* 52 (1): 35–55.

Stanovich, Keith E., and Richard F. West. 2000. "Individual Differences in Reasoning: Implications for the

Rationality Debate?" *Behavioral and Brain Sciences* 23 (5): 645–65.

Sydnor, Justin. 2010. "(Over)insuring Modest Risks." *American Economic Journal: Applied Economics* 2 (4): 177–99.

Taleb, Nassim Nicholas. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Thaler, Richard H., and Eric J. Johnson. 1990. "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice." *Management Science* 36 (6): 643–60.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven and London: Yale University Press.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–58.

Vissing-Jorgensen, Annette. 2004. "Perspectives on Behavioral Finance: Does 'Irrationality' Disappear with Wealth? Evidence from Expectations and Actions." In *NBER Macroeconomics Annual 2003*, edited by Mark Gertler and Kenneth Rogoff, 139–94. Cambridge and London: MIT Press.

Yagan, Danny. 2012. "Why Do Individual Investors Chase Stock Market Returns? Evidence of Belief in Long-Run Market Momentum." http://www.people.fas.harvard.edu/~yagan/papers/Chasing_Returns.pdf.

Maps of Bounded Rationality: Psychology for Behavioral Economics†

By DANIEL KAHNEMAN*

The work cited by the Nobel committee was done jointly with Amos Tversky (1937–1996) during a long and unusually close collaboration. Together, we explored the psychology of intuitive beliefs and choices and examined their bounded rationality. Herbert A. Simon (1955, 1979) had proposed much earlier that decision makers should be viewed as boundedly rational, and had offered a model in which utility maximization was replaced by satisficing. Our research attempted to obtain a map of bounded rationality, by exploring the systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational-agent models. The rational-agent model was our starting point and the main source of our null hypotheses, but Tversky and I viewed our research primarily as a contribution to psychology, with a possible contribution to economics as a secondary benefit. We were drawn into the interdisciplinary conversation by economists who hoped that psychology could be a useful source of assumptions for economic theorizing, and indirectly a source of hypotheses for economic research (Richard H. Thaler, 1980, 1991, 1992). These hopes have been realized to some extent, giving rise to an active program of research by behavioral economists (Thaler, 2000; Colin Camerer et al., forthcoming; for other examples, see Kahneman and Tversky, 2000).

My work with Tversky comprised three separate programs of research, some aspects of which were carried out with other collaborators. The first explored the heuristics that people use and the biases to which they are prone in various tasks of judgment under uncertainty, including predictions and evaluations of evidence (Kahneman and Tversky, 1973; Tversky and Kahneman, 1974; Kahneman et al., 1982). The second was concerned with prospect theory, a model of choice under risk (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) and with loss aversion in riskless choice (Kahneman et al., 1990, 1991; Tversky and Kahneman, 1991). The third line of research dealt with framing effects and with their implications for rational-agent models (Tversky and Kahneman, 1981, 1986). The present essay revisits these three lines of research in light of recent advances in the psychology of intuitive judgment and choice. Many of the ideas presented here were anticipated informally decades ago, but the attempt to integrate them into a coherent approach to judgment and choice is recent.

Economists often criticize psychological research for its propensity to generate lists of errors and biases, and for its failure to offer a coherent alternative to the rational-agent model. This complaint is only partly justified: psychological theories of intuitive thinking cannot match the elegance and precision of formal normative models of belief and choice, but this is just another way of saying that rational models are psychologically unrealistic. Furthermore, the alternative to simple and precise models is not chaos. Psychology offers integrative concepts and mid-level generalizations, which gain credibility from their ability to explain ostensibly different phenomena in diverse domains. In this spirit, the present essay offers a unified

† This article is a revised version of the lecture Daniel Kahneman delivered in Stockholm, Sweden, on December 8, 2002, when he received the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel. The article is copyright © The Nobel Foundation 2002 and is published here with the permission of the Nobel Foundation.

* Woodrow Wilson School, Princeton University, Princeton, NJ 08544 (e-mail: Kahneman@princeton.edu). This essay revisits problems that Amos Tversky and I studied together many years ago, and continued to discuss in a conversation that spanned several decades. It builds on an analysis of judgment heuristics that was developed in collaboration with Shane Frederick (Kahneman and Frederick, 2002). A different version was published in *American Psychologist* in September 2003. For detailed comments on this version I am grateful to Angus Deaton, David Laibson, Michael Rothschild, and Richard Thaler. The usual caveats apply. Geoffrey Goodwin, Amir Goren, and Kurt Schoppe provided helpful research assistance.

treatment of intuitive judgment and choice, which builds on an earlier study of the relationship between preferences and attitudes (Kahneman et al., 1999) and extends a model of judgment heuristics recently proposed by Kahneman and Shane Frederick (2002). The guiding ideas are (i) that most judgments and most choices are made intuitively; (ii) that the rules that govern intuition are generally similar to the rules of perception. Accordingly, the discussion of the rules of intuitive judgments and choices will rely extensively on visual analogies.

Section I introduces a distinction between two generic modes of cognitive function, corresponding roughly to intuition and reasoning. Section II describes the factors that determine the relative accessibility of different judgments and responses. Section III relates prospect theory to the general proposition that changes and differences are more accessible than absolute values. Section IV explains framing effects in terms of differential salience and accessibility. Section V reviews an attribute substitution model of heuristic judgment. Section VI describes a particular family of heuristics, called prototype heuristics. Section VII discusses the interactions between intuitive and deliberate thought. Section VIII concludes.

I. The Architecture of Cognition: Two Systems

The present treatment distinguishes two modes of thinking and deciding, which correspond roughly to the everyday concepts of reasoning and intuition. Reasoning is what we do when we compute the product of 17 by 258, fill an income tax form, or consult a map. Intuition is at work when we read the sentence "Bill Clinton is a shy man" as mildly amusing, or when we find ourselves reluctant to eat a piece of what we know to be chocolate that has been formed in the shape of a cockroach (Paul Rozin and Carol Nemeroff, 2002). Reasoning is done deliberately and effortfully, but intuitive thoughts seem to come spontaneously to mind, without conscious search or computation, and without effort. Casual observation and systematic research indicate that most thoughts and actions are normally intuitive in this sense (Daniel T. Gilbert, 1989, 2002; Timothy D. Wilson, 2002; Seymour Epstein, 2003).

Although effortless thought is the norm, some monitoring of the quality of mental operations and overt behavior also goes on. We do not express every passing thought or act on every impulse. But the monitoring is normally lax, and allows many intuitive judgments to be expressed, including some that are erroneous (Kahneman and Frederick, 2002). Ellen J. Langer et al. (1978) provided a well-known example of what she called "mindless behavior." In her experiment, a confederate tried to cut in line at a copying machine, using various preset "excuses." The conclusion was that statements that had the form of an unqualified request were rejected (e.g., "Excuse me, may I use the Xerox machine?"), but almost any statement that had the general form of an explanation was accepted, including "Excuse me, may I use the Xerox machine because I want to make copies?" The superficiality is striking.

Frederick (2003, personal communication) has used simple puzzles to study cognitive self-monitoring, as in the following example: "A bat and a ball cost \$1.10 in total. The bat costs \$1 more than the ball. How much does the ball cost?" Almost everyone reports an initial tendency to answer "10 cents" because the sum \$1.10 separates naturally into \$1 and 10 cents, and 10 cents is about the right magnitude. Frederick found that many intelligent people yield to this immediate impulse: 50 percent (47/93) of a group of Princeton students and 56 percent (164/293) of students at the University of Michigan gave the wrong answer. Clearly, these respondents offered their response without first checking it. The surprisingly high rate of errors in this easy problem illustrates how lightly the output of effortless associative thinking is monitored: people are not accustomed to thinking hard, and are often content to trust a plausible judgment that quickly comes to mind. Remarkably, Frederick has found that errors in this puzzle and in others of the same type were significant predictors of high discount rates.

In the examples discussed so far, intuition was associated with poor performance, but intuitive thinking can also be powerful and accurate. High skill is acquired by prolonged practice, and the performance of skills is rapid and effortless. The proverbial master chess player who walks past a game and declares "white mates in three" without slowing is performing intuitively (Simon and William G. Chase, 1973), as is the experienced nurse who

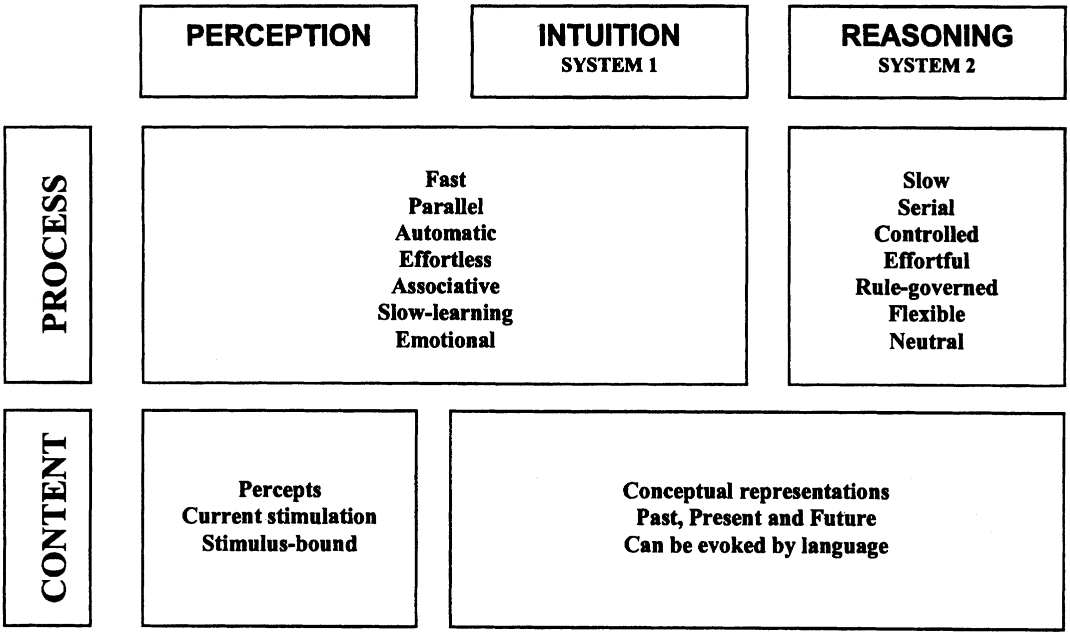


FIGURE 1. THREE COGNITIVE SYSTEMS

detects subtle signs of impending heart failure (Gary Klein, 1998; Atul Gawande, 2002).

The distinction between intuition and reasoning has recently been a topic of considerable interest to psychologists (see, e.g., Shelly Chaiken and Yaacov Trope, 1999; Gilbert, 2002; Steven A. Sloman, 2002; Keith E. Stanovich and Richard F. West, 2002). There is substantial agreement on the characteristics that distinguish the two types of cognitive processes, for which Stanovich and West (2000) proposed the neutral labels of System 1 and System 2. The scheme shown in Figure 1 summarizes these characteristics. The operations of System 1 are fast, automatic, effortless, associative, and often emotionally charged; they are also governed by habit, and are therefore difficult to control or modify. The operations of System 2 are slower, serial, effortful, and deliberately controlled; they are also relatively flexible and potentially rule-governed.

The difference in effort provides the most useful indications of whether a given mental process should be assigned to System 1 or System 2. Because the overall capacity for mental effort is limited, effortful processes tend to disrupt each other, whereas effortless processes neither cause nor suffer much interference when combined with other tasks. For example, a driver's ability to conduct a conversation is a sensitive indicator of the amount of attention currently demanded by the driving task. Dual tasks have been used in hundreds of psychological experiments to measure the attentional demands of different mental activities (for a review, see Harold E. Pashler, 1998). Studies using the dual-task method suggest that the self-monitoring function belongs with the effortful operations of System 2. People who are occupied by a demanding mental activity (e.g., attempting to hold in mind several digits) are much more likely to respond to another task by blurting out whatever comes to mind (Gilbert, 1989). The phrase that "System 2 monitors the activities of System 1" will be used here as shorthand for a hypothesis about what would happen if the operations of System 2 were disrupted. For example, it is safe to predict that the percentage of errors in the bat-and-ball question will increase, if the respondents are asked this question while attempting to keep a list of words in their active memory.

In the language that will be used here, the perceptual system and the intuitive operations of System 1 generate *impressions* of the attributes of objects of perception and thought. These impressions are not voluntary and need not be verbally explicit. In contrast, *judgments* are always explicit and intentional, whether or not they are overtly expressed. Thus, System 2 is involved in all judgments, whether they originate in impressions or in deliberate reasoning. The label "intuitive" is applied to judgments that directly reflect impressions.

Figure 1 illustrates an idea that guided the research that Tversky and I conducted from its early days: that intuitive judgments occupy a position—perhaps corresponding to evolutionary history—between the automatic operations of perception and the deliberate operations of reasoning. All the characteristics that students of intuition have attributed to System 1 are also properties of perceptual operations. Unlike perception, however, the operations of System 1 are not restricted to the processing of current stimulation. Like System 2, the operations of System 1 deal with stored concepts as well as with percepts, and can be evoked by language. This view of intuition suggests that the vast store of scientific knowledge available about perceptual phenomena can be a source of useful hypotheses about the workings of intuition. The strategy of drawing on analogies from perception is applied in the following section.

II. The Accessibility Dimension

A defining property of intuitive thoughts is that they come to mind spontaneously, like percepts. The technical term for the ease with which mental contents come to mind is *accessibility* (E. Tory Higgins, 1996). To understand intuition, we must understand why some thoughts are accessible and others are not. The remainder of this section introduces the concept of accessibility by examples drawn from visual perception.

Consider Figures 2a and 2b. As we look at the object in Figure 2a, we have immediate impressions of the height of the tower, the area of the top block, and perhaps the volume of the tower. Translating these impressions into units of height or volume requires a deliberate operation, but the impressions themselves are highly accessible. For other attributes, no perceptual impression exists. For example, the total area that the blocks would cover if the tower were dismantled is not perceptually accessible, though it can be estimated by a deliberate procedure, such as multiplying the area of a block by the number of blocks. Of course, the situation is reversed with Figure 2b. Now the blocks are laid out and an impression of total area is immediately accessible, but the height of the tower that could be constructed with these blocks is not.

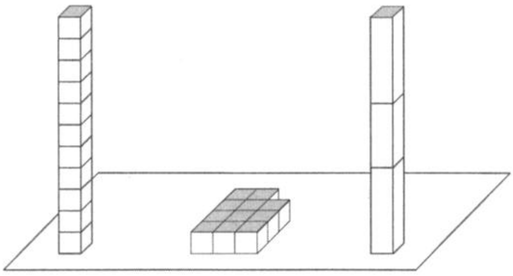


FIGURE 2. EXAMPLES OF DIFFERENTIAL ACCESSIBILITY

Some relational properties are accessible. Thus, it is obvious at a glance that Figures 2a and 2c are different, but also that they are more similar to each other than either is to Figure 2b. And some statistical properties of ensembles are accessible, while others are not. For an example, consider the question "What is the average length of the lines in Figure 3?" This question is easy. When a set of objects of the same general kind is presented to an observer—whether simultaneously or successively—a representation of the set is computed automatically, which includes quite precise information about the average (Dan Ariely, 2001; Sang-Chul Chong and Anne Treisman, 2003). The representation of the prototype is highly accessible, and it has the character of a percept: we form an impression of the typical line without choosing to do so. The only role for System 2 in this task is to map the impression of typical length onto the appropriate scale. In contrast, the answer to the question "What is the total length of the lines in the display?" does not come to mind without considerable effort.

As the example of averages and sums illustrates, some attributes are more accessible than others, both in perception and in judgment. Attributes that are routinely and automatically produced by the perceptual system or by System

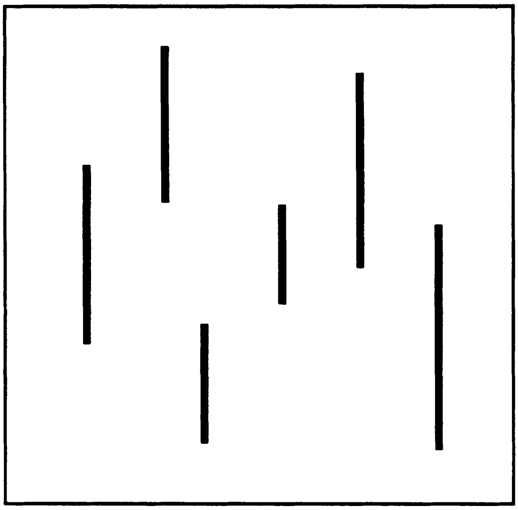


FIGURE 3. DIFFERENTIAL ACCESSIBILITY OF STATISTICAL PROPERTIES

1, without intention or effort, have been called *natural assessments* (Tversky and Kahneman, 1983). Kahneman and Frederick (2002) compiled a partial list of these natural assessments. In addition to physical properties such as size, distance, and loudness, the list includes more abstract properties such as similarity, causal propensity, surprisingness, affective valence, and mood.

The evaluation of stimuli as good or bad is a particularly important natural assessment. The evidence, both behavioral (John A. Bargh, 1997; Robert B. Zajonc, 1998) and neurophysiological (e.g., Joseph E. LeDoux, 2000), is consistent with the idea that the assessment of whether objects are good (and should be approached) or bad (should be avoided) is carried out quickly and efficiently by specialized neural circuitry. A remarkable experiment reported by Bargh (1997) illustrates the speed of the evaluation process, and its direct link to approach and avoidance. Participants were shown a series of stimuli on a screen, and instructed to respond to each stimulus as soon as it appeared, by moving a lever that blanked the screen. The stimuli were affectively charged words, some positive (e.g., LOVE) and some aversive (e.g., VOMIT), but this feature was irrelevant to the participant's task. Half the participants responded by pulling the lever toward themselves, half responded by pushing the lever away. Although the response was initiated within a fraction of a second, well before the meaning of the stimulus was consciously registered, the emotional valence of the word had a substantial effect. Participants were relatively faster in pulling a lever toward themselves (approach) for positive words, and relatively faster pushing the lever away when the word was aversive. The tendencies to approach or avoid were evoked by an automatic process that was not under conscious voluntary control. Several psychologists have commented on the influence of this primordial evaluative system (here included in System 1) on the attitudes and preferences that people adopt consciously and deliberately (Zajonc, 1998; Kahneman et al., 1999; Paul Slovic et al., 2002; Epstein, 2003).

The preceding discussion establishes a dimension of accessibility. At one end of this dimension we find operations that have the characteristics of perception and of the intuitive System 1: they are rapid, automatic, and effortless. At the other end are slow, serial, and effortful operations that people need a special reason to undertake. Accessibility is a continuum, not a dichotomy, and some effortful operations demand more effort than others. Some of the determinants of accessibility are probably genetic; others develop through experience. The acquisition of skill gradually increases the accessibility of useful responses and of productive ways to organize information, until skilled performance becomes almost effortless. This effect of practice is not limited to motor skills. A master chess player does not see the same board as the novice, and visualizing the tower in an array of blocks would also become virtually effortless with prolonged practice.

The impressions that become accessible in any particular situation are mainly determined, of course, by the actual properties of the object of judgment: it is easier to see a tower in Figure 2a than in Figure 2b, because the tower in the latter is only virtual. Physical salience also determines accessibility: if a large green letter and a small blue letter are shown at the same time, "green" will come to mind first. However, salience can be overcome by deliberate attention: an instruction to look for the small object will enhance the accessibility of all its features.

Analogous effects of salience and of spontaneous and voluntary attention occur with more abstract stimuli. For example, the statements "Team A beat team B" and "Team B lost to

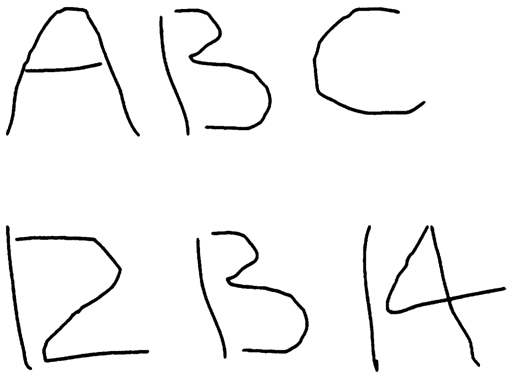


FIGURE 4. AN EFFECT OF CONTEXT ON ACCESSIBILITY

team A" convey the same information, but because each sentence draws attention to its grammatical subject, they make different thoughts accessible. Accessibility also reflects temporary states of associative activation. For example, the mention of a familiar social category temporarily increases the accessibility of the traits associated with the category stereotype, as indicated by a lowered threshold for recognizing behaviors as indications of these traits (Susan T. Fiske, 1998).

As designers of billboards know well, motivationally relevant and emotionally arousing stimuli spontaneously attract attention. Billboards are useful to advertisers because paying attention to an object makes all its features accessible—including those that are not linked to its primary motivational or emotional significance. The "hot" states of high emotional and motivational arousal greatly increase the accessibility of thoughts that relate to the immediate emotion and to the current needs, and reduce the accessibility of other thoughts (George Loewenstein, 1996, 2000; Jon Elster, 1998). An effect of emotional significance on accessibility was demonstrated in an important study by Yuval Rottenstreich and Christopher K. Hsee (2001), which showed that people are less sensitive to variations of probability when valuing chances to receive emotionally loaded outcomes (kisses and electric shocks) than when the outcomes are monetary.

Figure 4 (adapted from Jerome S. Bruner and A. Leigh Minturn, 1955) includes a standard demonstration of the effect of context on accessibility. An ambiguous stimulus that is perceived as a letter within a context of letters is instead seen as a number when placed within a context of numbers. More generally, expectations (conscious or not) are a powerful determinant of accessibility.

Another important point that Figure 4 illustrates is the complete suppression of ambiguity in conscious perception. This aspect of the demonstration is spoiled for the reader who sees the two versions in close proximity, but when the two lines are shown separately, observers will not spontaneously become aware of the alternative interpretation. They "see" the interpretation of the object that is the most likely in its context, but have no subjective indication that it could be seen differently. Ambiguity and uncertainty are suppressed in intuitive judgment as well as in perception. Doubt is a phenomenon of System 2, an awareness of one's ability to think incompatible thoughts about the same thing. The central finding in studies of intuitive decisions, as described by Klein (1998), is that experienced decision makers working under pressure (e.g., firefighting company captains) rarely need to choose between options because, in most cases, only a single option comes to mind.

The compound cognitive system that has been sketched here is an impressive computational device. It is well-adapted to its environment and has two ways of adjusting to changes: a short-term process that is flexible and effortful, and a long-term process of skill acquisition that eventually produces highly effective responses at low cost. The system tends to see what it expects to see—a form of Bayesian adaptation—and it is also capable of responding effectively to surprises. However, this marvelous creation differs in important respects from another paragon, the rational agent assumed in economic theory. Some of these differences are explored in the following sections, which review several familiar results as effects of accessibility. Possible implications for theorizing in behavioral economics are explored along the way.

III. Changes or States: Prospect Theory

A general property of perceptual systems is that they are designed to enhance the accessibility of changes and differences. Perception is *reference-dependent:* the perceived attributes of a focal stimulus reflect the contrast between that stimulus and a context of prior and concurrent stimuli. This section will show that

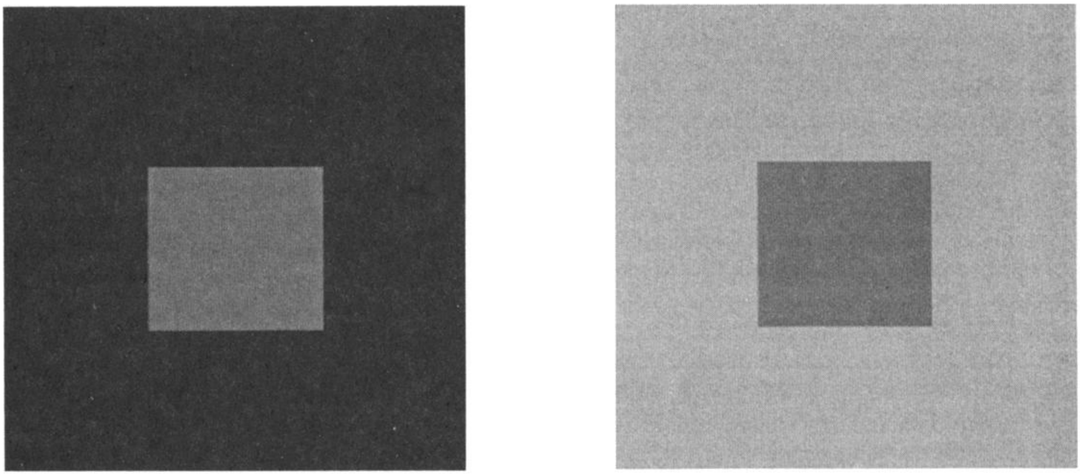


FIGURE 5. REFERENCE-DEPENDENCE IN THE PERCEPTION OF BRIGHTNESS

intuitive evaluations of outcomes are also reference-dependent.

The role of prior stimulation is familiar in the domain of temperature. Immersing the hand in water at 20°C will feel pleasantly warm after prolonged immersion in much colder water, and pleasantly cool after immersion in much warmer water. Figure 5 illustrates reference-dependence in vision. The two enclosed squares have the same luminance, but they do not appear equally bright. The point of the demonstration is that the brightness of an area is not a single-parameter function of the light energy that reaches the eye from that area, just as the experience of temperature is not a single-parameter function of the temperature to which one is currently exposed. An account of perceived brightness or temperature also requires a parameter for a reference value (often called adaptation level), which is influenced by the context of current and prior stimulation.

From the vantage point of a student of perception, it is quite surprising that in standard economic analyses the utility of decision outcomes is assumed to be determined entirely by the final state of endowment, and is therefore reference-independent. In the context of risky choice, this assumption can be traced to the brilliant essay that first defined a theory of expected utility (Daniel Bernoulli, 1738). Bernoulli assumed that states of wealth have a specified utility, and proposed that the decision rule for choice under risk is to maximize the expected utility of wealth (the moral expectation). The language of Bernoulli's essay is prescriptive—it speaks of what is sensible or reasonable to do—but the theory was also intended as a description of the choices of reasonable men (Gerd Gigerenzer et al., 1989). As in most modern treatments of decision-making, Bernoulli's essay does not acknowledge any tension between prescription and description. The proposition that decision makers evaluate outcomes by the utility of final asset positions has been retained in economic analyses for almost 300 years. This is rather remarkable, because the idea is easily shown to be wrong; I call it Bernoulli's error.

Tversky and I constructed numerous thought experiments when we began the study of risky choice that led to the formulation of prospect theory (Kahneman and Tversky, 1979). Examples such as Problems 1 and 2 below convinced us of the inadequacy of the utility function for wealth as an explanation of choice.

> Problem 1
> Would you accept this gamble?
>
> *50% chance to win $150*
> *50% chance to lose $100*
>
> Would your choice change if your overall wealth were lower by $100?

There will be few takers of the gamble in Problem 1. The experimental evidence shows that most people will reject a gamble with even chances to win and lose, unless the possible win is at least twice the size of the possible loss (e.g., Tversky and Kahneman, 1992). The answer to the second question is, of course, negative. Next consider Problem 2:

Problem 2
Which would you choose?
lose $100 with certainty
or
50% chance to win $50
50% chance to lose $200

Would your choice change if your overall wealth were higher by $100?

In Problem 2, the gamble appears much more attractive than the sure loss. Experimental results indicate that risk-seeking preferences are held by a large majority of respondents in problems of this kind (Kahneman and Tversky, 1979). Here again, the idea that a change of $100 in total wealth would affect preferences cannot be taken seriously.

We examined many choice pairs of this type in our early explorations, and concluded that the very abrupt switch from risk aversion to risk seeking could not plausibly be explained by a utility function for wealth. Preferences appeared to be determined by attitudes to gains and losses, defined relative to a reference point, but Bernoulli's theory and its successors did not incorporate a reference point. We therefore proposed an alternative theory of risk, in which the carriers of utility are gains and losses—changes of wealth rather than states of wealth. One novelty of prospect theory was that it was explicitly presented as a formal descriptive theory of the choices that people actually make, not as a normative model. This was a departure from a long history of choice models that served double duty as normative logics and as idealized descriptive models.

The distinctive predictions of prospect theory follow from the shape of the value function, which is shown in Figure 6. The value function is defined on gains and losses and is characterized by three features: (1) it is concave in the domain of gains, favoring risk aversion; (2) it is convex in the domain of losses, favoring risk seeking; (3) most important, the function is sharply kinked at the reference point, and *loss-averse*—steeper for losses than for gains by a factor of about 2–2.5 (Kahneman et al., 1991; Tversky and Kahneman, 1992).

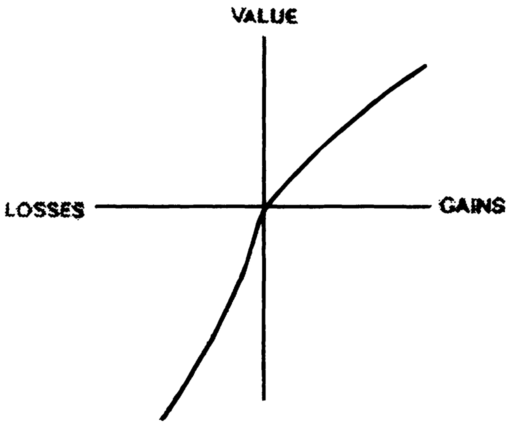


FIGURE 6. A SCHEMATIC VALUE FUNCTION FOR CHANGES

If Bernoulli's formulation is transparently incorrect as a descriptive model of risky choices, as has been argued here, why has this model been retained for so long? The answer appears to be that the assignment of utility to wealth is an aspect of rationality, and therefore compatible with the general assumption of rationality in economic theorizing (Kahneman, 2003a). Consider Problem 3:

Problem 3
Two persons get their monthly report from a broker:
A is told that her wealth went from 4M to 3M
B is told that her wealth went from 1M to 1.1M

Who of the two individuals has more reason to be satisfied with her financial situation?

Who is happier today?

Problem 3 highlights the contrasting interpretations of utility in theories that define outcomes as states or as changes. In Bernoulli's analysis only the first of the two questions of Problem 3 is relevant, and only long-term consequences matter. Prospect theory, in contrast, is concerned with short-term outcomes, and the value function presumably reflects an anticipation of the valence and intensity of the emotions that will be experienced at moments of transition from one state to another (Kahneman, 2000a, b; Barbara Mellers, 2000). Which of these concepts of utility is more useful? The cultural norm of reasonable decision-making favors the long-term view over a concern with transient emotions. Indeed, the adoption of a broad perspective and a long-term view is an aspect of the meaning of rationality in everyday language. The final-states interpretation of the utility of outcomes is therefore a good fit for a rational-agent model.

These considerations support the normative and prescriptive status of the Bernoullian definition of outcomes. On the other hand, an exclusive concern with the long term may be prescriptively sterile, because the long term is not where life is lived. Utility cannot be divorced from emotion, and emotions are triggered by changes. A theory of choice that completely ignores feelings such as the pain of losses and the regret of mistakes is not only descriptively unrealistic, it also leads to prescriptions that do not maximize the utility of outcomes as they are actually experienced—that is, utility as Bentham conceived it (Kahneman, 1994, 2000a; Kahneman et al., 1997).

Bernoulli's error—the idea that the carriers of utility are final states—is not restricted to decision-making under risk. Indeed, the incorrect assumption that initial endowments do not matter is the basis of Coase's theorem and of its multiple applications (Kahneman et al., 1990). The error of reference-independence is built into the standard representation of indifference maps. It is puzzling to a psychologist that these maps do not include a representation of the decision maker's current holdings of various goods—the counterpart of the reference point in prospect theory. The parameter is not included, of course, because consumer theory assumes that it does not matter.

The core idea of prospect theory—that the value function is kinked at the reference point and loss averse—became useful to economics when Thaler (1980) used it to explain riskless choices. In particular, loss aversion explained a violation of consumer theory that Thaler identified and labeled the "endowment effect": the selling price for consumption goods is much higher than the buying price, often by a factor of 2 or more. The value of a good to an individual appears to be higher when the good is viewed as something that could be lost or given up than when the same good is evaluated as a potential gain (Kahneman et al., 1990, 1991; Tversky and Kahneman, 1991).

When half the participants in an experimental market were randomly chosen to be endowed with a good (a mug) and trade was allowed, the volume of trade was about half the amount that would be predicted by assuming that value was independent of initial endowment (Kahneman et al., 1990). Transaction costs did not explain this counterexample to the Coase theorem, because the same institution produced no indication of reluctance to trade when the objects of trade were money tokens. The results suggest that the participants in these experiments did not value the mug as an object they could have and consume, but as something they could get, or give up. Interestingly, John A. List (2003a, b) found that the magnitude of the endowment effect was substantially reduced for participants with intense experience in the trading of sportscards. Experienced traders (who are also consumers) showed less reluctance to trade one good for another—not only sportscards, but also mugs and other goods—as if they had learned to base their choice on long-term value, rather than on the immediate emotions associated with getting or giving up objects.

Reference-dependence and loss aversion help account for several phenomena of choice. The familiar observation that out-of-pocket losses are valued much more than opportunity costs is readily explained, if these outcomes are evaluated on different limbs of the value function. The distinction between "actual" losses and losses of opportunities is recognized in many ways in the law (David Cohen and Jack L. Knetsch, 1992) and in lay intuitions about rules of fairness in the market (Kahneman et al., 1986). Loss aversion also contributes to the well-documented status-quo bias (William Samuelson and Richard Zeckhauser, 1988). Because the reference point is usually the status quo, the properties of alternative options are evaluated as advantages or disadvantages relative to the current situation, and the disadvantages of the alternatives loom larger than their advantages. Other applications of the concept of loss aversion are documented in several chapters in Kahneman and Tversky (2000).

IV. Framing Effects

In the display of blocks in Figure 2, the same property (the total height of a set of blocks) was highly accessible in one display and not in another, although both displays contained the same information. This observation is entirely unremarkable—it does not seem shocking that some attributes of a stimulus are automatically perceived while others must be computed, or that the same attribute is perceived in one display of an object but must be computed in another. In the context of decision-making, however, similar observations raise a significant challenge to the rational-agent model.

The assumption that preferences are not affected by inconsequential variations in the description of outcomes has been called extensionality (Kenneth J. Arrow, 1982) and invariance (Tversky and Kahneman, 1986), and is considered an essential aspect of rationality. Invariance is violated in *framing effects,* where extensionally equivalent descriptions lead to different choices by altering the relative salience of different aspects of the problem. Tversky and Kahneman (1981) introduced their discussion of framing effects with the following problem:

> The Asian disease
>
> *Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:*
>
> *If Program A is adopted, 200 people will be saved*
>
> *If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved*

In this version of the problem, a substantial majority of respondents favor Program A, indicating risk aversion. Other respondents, selected at random, receive a question in which the same cover story is followed by a different description of the options:

> *If Program A′ is adopted, 400 people will die*
>
> *If Program B′ is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die*

A substantial majority of respondents now favor Program B′, the risk-seeking option. Although there is no substantive difference between the versions, they evoke different associations and evaluations. This is easiest to see in the certain option, because outcomes that are certain are overweighted relative to outcomes of high or intermediate probability (Kahneman and Tversky, 1979). Thus, the certainty of saving people is disproportionately attractive, while accepting the certain death of people is disproportionately aversive. These immediate affective responses respectively favor A over B and B′ over A′. As in Figures 2a and 2b, the different representations of the outcomes highlight some features of the situation and mask others.

In an essay about the ethics of policy, Thomas C. Schelling (1984) presented a compellingly realistic example of the dilemmas raised by framing. Schelling reports asking his students to evaluate a tax policy that would allow a larger child exemption to the rich than to the poor. Not surprisingly, his students found this proposal outrageous. Schelling then pointed out that the default case in the standard tax table is a childless family, with special adjustments for families with children, and led his class to agree that the existing tax schedule could be rewritten with a family with two children as the default case. In this formulation, childless families would pay a surcharge. Should this surcharge be as large for the poor as for the rich? Of course not. The two versions of the question about how to treat the rich and the poor both trigger an intuitive preference for protecting the poor, but these preferences are incoherent. Schelling's problem highlights an important point. Framing effects are not a laboratory curiosity, but a ubiquitous reality. The tax table must be framed one way or another, and each frame will increase the accessibility of some responses and make other responses less likely.

There has been considerable interest among behavioral economists in a particular type of framing effect, where a choice between two options A and B is affected by designating either A or B as a default option. The option designated as the default has a large advantage in such choices, even for decisions that have considerable significance. Eric J. Johnson et al. (1993) described a compelling example. The states of Pennsylvania and New Jersey both offer drivers a choice between an insurance policy that allows an unconstrained right to sue, and a less expensive policy that restricts the right to sue. The unconstrained right to sue is the default in Pennsylvania, the opposite is the default in New Jersey, and the takeup of full coverage is 79 percent and 30 percent in the two states, respectively. Johnson and Daniel G. Goldstein (2003) estimate that Pennsylvania drivers spend 450 million dollars annually on full coverage that they would not purchase if their choice were framed as it is for New Jersey drivers.

Johnson and Goldstein (2003) also compared the proportions of the population enrolled in organ donation programs in seven European countries in which enrollment was the default and four in which nonenrollment was the default. Averaging over countries, enrollment in donor programs was 97.4 percent when this was the default option, 18 percent otherwise. The passive acceptance of the formulation given has significant consequences in this case, as it does in other recent studies where the selection of the default on the form that workers completed to set their 401(k) contributions dominated their ultimate choice (Brigitte Madrian and Dennis Shea, 2001; James J. Choi et al., 2002).

The basic principle of framing is the passive acceptance of the formulation given. Because of this passivity, people fail to construct a canonical representation for all extensionally equivalent descriptions of a state of affairs. They do not spontaneously compute the height of a tower that could be built from an array of blocks, and they do not spontaneously transform the representation of puzzles or decision problems. Obviously, no one is able to recognize "137 × 24" and "3,288" as "the same" number without going through some elaborate computations. Invariance cannot be achieved by a finite mind.

The impossibility of invariance raises significant doubts about the descriptive realism of rational-choice models (Tversky and Kahneman, 1986). Absent a system that reliably generates appropriate canonical representations, intuitive decisions will be shaped by the factors that determine the accessibility of different features of the situation. Highly accessible features will influence decisions, while features of low accessibility will be largely ignored—and the correlation between accessibility and reflective judgments of relevance in a state of complete information is not necessarily high.

A particularly unrealistic assumption of the rational-agent model is that agents make their choices in a comprehensively inclusive context, which incorporates all the relevant details of the present situation, as well as expectations about all future opportunities and risks. Much evidence supports the contrasting claim that people's views of decisions and outcomes are normally characterized by "narrow framing" (Kahneman and Daniel Lovallo, 1993), and by the related notions of "mental accounting" (Thaler, 1985, 1999) and "decision bracketing" (Daniel Read et al., 1999).

The following are some examples of the prevalence of narrow framing. The decision of whether or not to accept a gamble is normally considered as a response to a single opportunity, not as an occasion to apply a general policy (Gideon Keren and Willem A. Wagenaar, 1987; Tversky and Donald A. Redelmeier, 1992; Kahneman and Lovallo, 1993; Shlomo Benartzi and Thaler, 1999). Investors' decisions about particular investments appear to be considered in isolation from the remainder of the investor's portfolio (Nicholas Barberis et al., 2003). The time horizon that investors adopt for evaluating their investments appears to be unreasonably short—an observation that helps explain the equity-premium puzzle (Benartzi and Thaler, 1995). Finally, the prevalence of the gain/loss framing of outcomes over the wealth frame, which was discussed in the previous section, can now be seen as an instance of narrow framing. A shared feature of all these examples is that decisions made in narrow frames depart far more from risk neutrality than decisions that are made in a more inclusive context.

The prevalence of narrow frames is an effect of accessibility, which can be understood by referring to the displays of blocks in Figure 2. The same set of blocks is framed as a tower in Figure 2a, and as a flat array in Figure 2b. Although it is possible to "see" a tower in Figure 2b, it is much easier to do so in Figure 2a. Narrow frames generally reflect the structure of the environment in which decisions are made. The choices that people face arise one at a time, and the principle of passive acceptance suggests that they will be considered as they arise. The problem at hand and the immediate consequences of the choice will be far more accessible than all other considerations, and as a result decision problems will be framed far more narrowly than the rational model assumes.

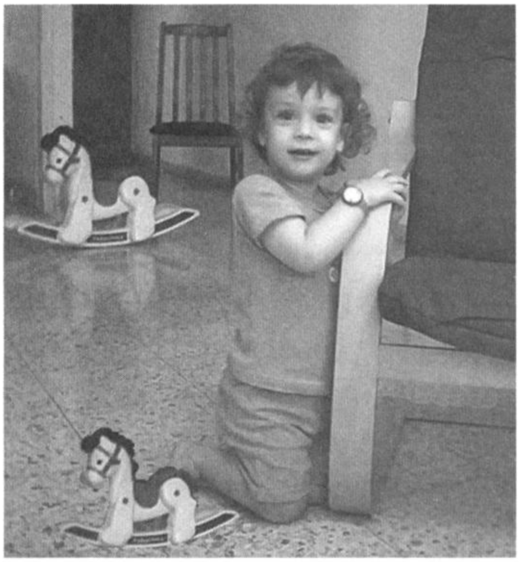


FIGURE 7. AN ILLUSION OF ATTRIBUTE SUBSTITUTION

Source: Photo by Lenore Shoham, 2003.

V. Attribute Substitution: A Model of Judgment Heuristics

The first research program that Tversky and I undertook together consisted of a series of studies of various types of judgment about uncertain events, including numerical predictions and assessments of the probabilities of hypotheses. Our conclusion in a review of this work was that "people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors" (Tversky and Kahneman, 1974, p. 1124). The article introduced three heuristics—representativeness, availability, and anchoring—that were used to explain a dozen systematic biases in judgment under uncertainty, including nonregressive prediction, neglect of base-rate information, overconfidence, and overestimates of the frequency of events that are easy to recall. Some of the biases were identified by systematic errors in estimates of known quantities and statistical facts. Other biases were defined by discrepancies between the regularities of intuitive judgments and the principles of probability theory, Bayesian inference, and regression analysis.

Kahneman and Frederick (2002) recently revisited the early studies of judgment heuristics, and proposed a formulation in which the reduction of complex tasks to simpler operations is achieved by an operation of *attribute substitution.* "Judgment is said to be mediated by a heuristic when the individual assesses a specified *target attribute* of a judgment object by substituting another property of that object—the *heuristic attribute*—which comes more readily to mind" (p. 53). Unlike the early work, Kahneman and Frederick's conception of heuristics is not restricted to the domain of judgment under uncertainty.

For a perceptual example of attribute substitution, consider the question: "What are the sizes of the two horses in Figure 7, as they are drawn on the page?" The images are in fact identical in size, but the figure produces a compelling illusion. The target attribute that observers intend to evaluate is objective two-dimensional size, but they are unable to do this veridically. Their judgments map an impression of three-dimensional size (the heuristic attribute) onto units of length that are appropriate to the target attribute, and scaled to the size of the page. This illusion is caused by the differential accessibility of competing interpretations of the image. An impression of three-represented by prototypes. All these features of the cognitive system were in our minds in some form when Amos Tversky and I began our joint work in 1969, and most of them were in Herbert Simon's mind much earlier. However, the role of emotion in judgment and decision making received less attention in that work than it had received before the beginning of the cognitive revolution in psychology in the 1950's. More recent developments have restored a central role to emotion, which is incorporated in the view of intuition that was presented here. Findings about the role of optimism in risk taking, the effects of emotion on decision weights, the role of fear in predictions of harm, and the role of liking and disliking in factual predictions—all indicate that the traditional separation between belief and preference in analyses of decision making is psychologically unrealistic.

Incorporating a common sense psychology of the intuitive agent into economic models will present difficult challenges, especially for formal theorists. It is encouraging to note, however, that the challenge of incorporating the first wave of psychological findings into economics appeared even more daunting 20 years ago, and that challenge has been met with considerable success.

REFERENCES

Agnoli, Franca. "Development of Judgmental Heuristics and Logical Reasoning: Training Counteracts the Representativeness Heuristic." *Cognitive Development*, April–June 1991, *6*(2), pp. 195–217.

Agnoli, Franca and Krantz, David H. "Suppressing Natural Heuristics by Formal Instruction: The Case of the Conjunction Fallacy." *Cognitive Psychology*, October 1989, *21*(4), pp. 515–50.

Alevy, Jonathan E.; List, John A. and Adamowicz, Wiktor. "More is Less: Preference Reversals and Non-Market Valuations." Working paper, University of Maryland, 2003.

Ariely, Dan. "Seeing Sets: Representation by Statistical Properties." *Psychological Science*, March 2001, *12*(2), pp. 157–62.

Ariely, Dan and Loewenstein, George. "When Does Duration Matter in Judgment and Decision Making?" *Journal of Experimental Psychology: General*, December 2000, *129*(4), pp. 508–23.

Arrow, Kenneth J. "Risk Perception in Psychology and Economics." *Economic Inquiry*, January 1982, *20*(1), pp. 1–9.

Barber, Brad M. and Odean, Terrance. "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance*, April 2000, *55*(2), pp. 773–806.

Barberis, Nicholas; Huang, Ming and Thaler, Richard H. "Individual Preferences, Monetary Gambles and the Equity Premium." National Bureau of Economic Research (Cambridge, MA) Working Paper No. W9997, May 2003.

Bargh, John A. "The Automaticity of Everyday Life," in Robert S. Wyer, Jr., ed., *The automaticity of everyday life: Advances in social cognition*, Vol. 10. Mahwah, NJ: Erlbaum, 1997, pp. 1–61.

Benartzi, Shlomo and Thaler, Richard H. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, February 1995, *110*(1), pp. 73–92.

______. "Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments." *Management Science*, March 1999, *47*(3), pp. 364–81.

______. "Naïve Diversification Strategies in Defined Contribution Saving Plans." *American Economic Review*, March 2001, *91*(1), pp. 79–98.

Bernoulli, Daniel. "Exposition of a New Theory on the Measurement of Risk." *Econometrica*, January 1954, *22*(1), pp. 23–36. (Original work published 1738.)

Bless, Herbert; Clore, Gerald L.; Schwarz, Norbert; Golisano, Verana; Rabe, Christian and Wolk, Marcus. "Mood and the Use of Scripts: Does a Happy Mood Really Lead to Mindlessness?" *Journal of Personality and Social Psychology*, October 1996, *71*(4), pp. 665–79.

Bodenhausen, Galen V. "Stereotypes as Judgmental Heuristics: Evidence of Circadian Variations in Discrimination." *Psychological Science*, September 1990, *1*(5), pp. 319–22.

Bruner, Jerome S. and Minturn, A. Leigh. "Perceptual Identification and Perceptual Organization." *Journal of General Psychology*, July 1955, *53*, pp. 21–28.

Camerer, Colin F. and Hogarth, Robin M. "The Effect of Financial Incentives." *Journal of Risk and Uncertainty*, December 1999, *19*(1–3), pp. 7–42.

Camerer, Colin F.; Loewenstein, George and Rabin, Matthew, eds. *Advances in behavioral economics*. Princeton, NJ: Princeton University Press (forthcoming).

Carson, Richard T. "Contingent Valuation Surveys and Tests of Insensitivity to Scope," in R. J. Kopp, W. W. Pommerhene, and N. Schwartz, eds., *Determining the value of non-marketed goods: Economic, psychological, and policy relevant aspects of contingent valuation methods*. Boston, MA: Kluwer, 1997, pp. 127–63.

Chaiken, Shelly and Trope, Yaacov, eds. *Dual-process theories in social psychology*. New York: Guilford Press, 1999.

Chapman, Gretchen B. and Johnson, Eric J. "Incorporating the Irrelevant: Anchors in Judgments of Belief and Value," in Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds., *Heuristics and biases: The psychology of intuitive thought*. New York: Cambridge University Press, 2002, pp. 120–38.

Choi, James J.; Laibson, David; Madrian, Brigitte and Metrick, Andrew. "Defined Contribution Pensions: Plan Rules, Participant Decisions and the Path of Least Resistance," in James M. Poterba, ed., *Tax policy and the economy*, Vol. 16. Cambridge, MA: MIT Press, 2002, pp. 67–113.

Chong, Sang-Chul and Treisman, Anne. "Representation of Statistical Properties." *Vision Research*, February 2003, *43*(4), pp. 393–404.

Cohen, David and Knetsch, Jack L. "Judicial Choice and Disparities Between Measures of Economic Value." *Osgoode Hall Law Review*, 1992, *30*(3), pp. 737–70.

Cosmides, Leda and Tooby, John. "Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions From the Literature on Judgment and Uncertainty." *Cognition*, January 1996, *58*(1), pp. 1–73.

De Bondt, Werner F. M. and Thaler, Richard H. "Does the Stock Market Overreact?" *Journal of Finance*, July 1985, *40*(3), pp. 793–808.

Desvousges, William H.; Johnson, F. Reed; Dunford, Richard W.; Hudson, Sara P.; Wilson, K. Nichole and Boyle, Kevin J. "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability," in Jerry A. Hausman, ed., *Contingent valuation: A critical assessment*. Amsterdam: North-Holland, 1993, pp. 91–164.

Diamond, Peter A. "A Framework for Social Security Analysis." *Journal of Public Economics*, December 1977, *8*(3), pp. 275–98.

______. "Testing the Internal Consistency of Contingent Valuation Surveys." *Journal of Environmental Economics and Management*, May 1996, *30*(3), pp. 155–73.

Elster, Jon. "Emotions and Economic Theory." *Journal of Economic Literature*, March 1998, *26*(1), pp. 47–74.

Epstein, Seymour. "Cognitive-Experiential Self-Theory of Personality," in Theodore Millon and Melvin J. Lerner, eds., *Comprehensive handbook of psychology, volume 5: Personality and social psychology*. Hoboken, NJ: Wiley & Sons, 2003, pp. 159–84.

Evans, Jonathan St. B. T.; Handley, Simon J.; Over, David E. and Perham, Nicholas. "Background Beliefs in Bayesian Inference." *Memory and Cognition*, March 2002, *30*(2), pp. 179–90.

Finucane, Melissa L.; Alhakami, Ali; Slovic, Paul and Johnson, Stephen M. "The Affect Heuristic in Judgments of Risks and Benefits." *Journal of Behavioral Decision Making*, January/March 2000, *13*(1), pp. 1–17.

Fiske, Susan T. "Stereotyping, Prejudice, and Discrimination," in Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, eds., *The handbook of social psychology*, 4th Ed., Vol. 1. New York: McGraw-Hill, 1998, pp. 357–411.

Frederick, Shane W. and Fischhoff, Baruch. "Scope (In)sensitivity in Elicited Valuations." *Risk, Decision, and Policy*, August 1998, *3*(2), pp. 109–23.

Fredrickson, Barbara L. and Kahneman, Daniel. "Duration Neglect in Retrospective Evaluations of Affective Episodes." *Journal of Personality and Social Psychology*, July 1993, *65*(1), pp. 45–55.

Gawande, Atul. *Complications: A surgeon's notes on an imperfect science*. New York: Metropolitan Books, 2002.

Genesove, David and Mayer, Christopher J. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *Quarterly Journal of Economics*, November 2001, *116*(4), pp. 1233–60.

Gigerenzer, Gerd; Swijtink, Zeno; Porter, Theodore; Daston, Lorraine; Beatty, John and Kruger, Lorenz. *The empire of chance: How probability changed science and everyday* *life*. Cambridge: Cambridge University Press, 1989.

Gilbert, Daniel T. "Thinking Lightly About Others: Automatic Components of the Social Inference Process," in James S. Uleman and John A. Bargh, eds., *Unintended thought*. Englewood Cliffs, NJ: Prentice-Hall, 1989, pp. 189–211.

______. "How Mental Systems Believe." *American Psychologist*, February 1991, *46*(2), pp. 107–19.

______. "Inferential Correction," in Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds., *Heuristics and biases: The psychology of intuitive thought*. New York: Cambridge University Press, 2002, pp. 167–84.

Grether, David M. "Recent Psychological Studies of Behavior Under Uncertainty." *American Economic Review*, May 1978 (*Papers and Proceedings*), *68*(2), pp. 70–74.

Higgins, E. Tory. "Knowledge Activation: Accessibility, Applicability, and Salience," in E. Tory Higgins and Arie W. Kruglanski, eds., *Social psychology: Handbook of basic principles*. New York: Guilford Press, 1996, pp. 133–68.

Hsee, Christopher K. "The Evaluability Hypothesis: An Explanation of Preference Reversals Between Joint and Separate Evaluations of Alternatives." *Organizational Behavior and Human Decision Processes*, September 1996, *67*(3), pp. 247–57.

______. "Less is Better: When Low-Value Options are Valued More Highly Than High-Value Options." *Journal of Behavioral Decision Making*, June 1998, *11*(2), pp. 107–21.

Isen, Alice M.; Nygren, Thomas E. and Ashby, F. Gregory. "Influence of Positive Affect on the Subjective Utility of Gains and Losses: It is Just Not Worth the Risk." *Journal of Personality and Social Psychology*, November 1988, *55*(5), pp. 710–17.

Johnson, Eric J. and Goldstein, Daniel G. "Do Defaults Save Lives?" Working paper, Center for Decision Sciences, Columbia University, 2003.

Johnson, Eric J.; Hershey, John; Meszaros, Jacqueline and Kunreuther, Howard. "Framing, Probability Distortions, and Insurance Decisions." *Journal of Risk and Uncertainty*, August 1993, *7*(1), pp. 35–51.

Kahneman, Daniel. "Comment," in Ronald G. Cummings, David S. Brookshire, and William D. Schultze, eds., *Valuing environmental goods*. Totowa, NJ: Rowman and Allenheld, 1986, pp. 185–93.

______. "New Challenges to the Rationality Assumption." *Journal of Institutional and Theoretical Economics*, March 1994, *150*(1), pp. 18–36.

______. "Evaluation by Moments: Past and Future," in Daniel Kahneman and Amos Tversky, eds., *Choices, values, and frames*. New York: Cambridge University Press, 2000a, pp. 693–708.

______. "Experienced Utility and Objective Happiness: A Moment-Based Approach," in Daniel Kahneman and Amos Tversky, eds., *Choices, values, and frames*. New York: Cambridge University Press, 2000b, pp. 673–92.

______. "A Psychological Perspective on Economics." *American Economic Review*, May 2003a (*Papers and Proceedings*), *93*(2), pp. 162–68.

______. "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist*, September 2003b, *56*(9), pp. 697–720.

Kahneman, Daniel and Frederick, Shane. "Representativeness Revisited: Attribute Substitution in Intuitive Judgment," in Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds., *Heuristics and biases: The psychology of intuitive thought*. New York: Cambridge University Press, 2002, pp. 49–81.

Kahneman, Daniel; Fredrickson, Barbara L.; Schreiber, Charles A. and Redelmeier, Donald A. "When More Pain is Preferred to Less: Adding a Better End." *Psychological Science*, November 1993, *4*(6), pp. 401–05.

Kahneman, Daniel; Knetsch, Jack and Thaler, Richard. "Fairness as a Constraint on Profit-seeking: Entitlements in the Market." *American Economic Review*, September 1986, *76*(4), pp. 728–41.

______. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, December 1990, *98*(6), pp. 1325–48.

______. "The Endowment Effect, Loss Aversion, and Status Quo Bias: Anomalies." *Journal of Economic Perspectives*, Winter 1991, *5*(1), pp. 193–206.

Kahneman, Daniel and Lovallo, Daniel. "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." *Management Science*, January 1993, *39*(1), pp. 17–31.

Kahneman, Daniel; Ritov, Ilana and Schkade, David. "Economic Preferences or Attitude Expressions? An Analysis of Dollar Responses to Public Issues." *Journal of Risk and Uncertainty*, December 1999, *19*(1–3), pp. 203–35.

Kahneman, Daniel; Slovic, Paul and Tversky, Amos, eds. *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press, 1982.

Kahneman, Daniel and Tversky, Amos. "On the Psychology of Prediction." *Psychological Review*, July 1973, *80*(4), pp. 237–51.

______. "Prospect Theory: An Analysis of Decisions Under Risk." *Econometrica*, March 1979, *47*(2), pp. 263–91.

______, eds. *Choices, values, and frames*. New York: Cambridge University Press, 2000.

Kahneman, Daniel; Wakker, Peter P. and Sarin, Rakesh. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics*, May 1997, *112*(2), pp. 375–405.

Keren, Gideon and Wagenaar, Willem A. "Violations of Utility Theory in Unique and Repeated Gambles." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, July 1987, *13*(3), pp. 387–91.

Klein, Gary. *Sources of power: How people make decisions*. Cambridge, MA: MIT Press, 1998.

______. *Intuition at work: Why developing your gut instincts will make you better at what you do*. New York: Doubleday, 2003.

Koehler, Jonathan Jay. "The Base-Rate Fallacy Reconsidered: Descriptive, Normative, and Methodological Challenges." *Behavioral and Brain Sciences*, March 1996, *19*, pp. 1–53.

Kopp, Raymond. "Why Existence Value Should be Used in Cost-Benefit Analysis." *Journal of Policy Analysis and Management*, Winter 1992, *11*(1), pp. 123–30.

Kunreuther, Howard. "The Changing Societal Consequences of Risks From Natural Hazards." *Annals of the American Academy of Political and Social Science*, May 1979, *443*(443), pp. 104–16.

Langer, Ellen J.; Blank, Arthur and Chanowitz, Benzion. "The Mindlessness of Ostensibly Thoughtful Action: The Role of 'Placebic' Information in Interpersonal Interaction." *Journal of Personality and Social Psychology*, June 1978, *36*(6), pp. 635–42.

LeDoux, Joseph E. "Emotion Circuits in the Brain." *Annual Review of Neuroscience*, March 2000, *23*, pp. 155–84.

Lerner, Jennifer S. and Tetlock, Philip E. "Accounting for the Effects of Accountability." *Psychological Bulletin*, March 1999, *125*(2), pp. 255–75.

List, John A. "Preference Reversals of a Different Kind: The 'More Is Less' Phenomenon." *American Economic Review*, December 2002, *92*(5), pp. 1636–43.

______. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, February 2003a, *118*(1), pp. 47–71.

______. "Neoclassical Theory Versus Prospect Theory: Evidence From the Marketplace." National Bureau of Economic Research (Cambridge, MA) Working Paper No. W9736, 2003b; *Econometrica*, 2004 (forthcoming).

Loewenstein, George. "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, March 1996, *65*(3), pp. 272–92.

______. "Emotions in Economic Theory and Economic Behavior." *American Economic Review*, May 2000 (*Papers and Proceedings*), *90*(2), pp. 426–32.

Loewenstein, George; Weber, Elke U.; Hsee, Christopher K. and Welch, N. "Risk as Feelings." *Psychological Bulletin*, March 2001, *127*(2), pp. 267–86.

Luce, R. Duncan; Krantz, David H.; Suppes, Patrick and Tversky, Amos. *Foundations of measurement, volume 3: Representation, axiomatization, and invariance*. San Diego, CA: Academic Press, 1990.

Madrian, Brigitte and Shea, Dennis. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics*, November 2001, *116*(4), pp. 1149–87.

Mellers, Barbara. "Choice and the Relative Pleasure of Consequences." *Psychological Bulletin*, November 2000, *126*(6), pp. 910–24.

Nisbett, Richard E.; Krantz, David H.; Jepson, Christopher and Kunda, Ziva. "The Use of Statistical Heuristics in Everyday Inductive

Reasoning." *Psychological Review*, October 1983, *90*(4), pp. 339–63.

Pashler, Harold E. *The psychology of attention.* Cambridge, MA: MIT Press, 1998.

Posner, Michael I. and Keele, Stephen W. "On the Genesis of Abstract Ideas." *Journal of Experimental Psychology*, Pt. 1, 1968, *77*(3), pp. 353–63.

Rabin, Matthew. "Inference by Believers in the Law of Small Numbers." *Quarterly Journal of Economics*, August 2002, *17*(3), pp. 775–816.

Read, Daniel; Loewenstein, George and Rabin, Matthew. "Choice Bracketing." *Journal of Risk and Uncertainty*, December 1999, *19*(1–3), pp. 171–97.

Redelmeier, Donald A. and Kahneman, Daniel. "Patients' Memories of Painful Medical Treatments: Real-time and Retrospective Evaluations of Two Minimally Invasive Procedures." *Pain*, July 1996, *66*(1), pp. 3–8.

Redelmeier, Donald A.; Katz, Joel and Kahneman, Daniel. "Memories of Colonoscopy: A Randomized Trial." *Pain*, July 2003, *104*(1–2), pp. 187–94.

Rosch, Eleanor and Mervis, Carolyn B. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology*, October 1975, *7*(4), pp. 573–605.

Ross, Lee and Nisbett, Richard E. *The person and the situation.* New York: McGraw-Hill, 1991.

Rottenstreich, Yuval and Hsee, Christopher K. "Money, Kisses and Electric Shocks: On the Affective Psychology of Risk." *Psychological Science*, May 2001, *12*(3), pp. 185–90.

Rozin, Paul and Nemeroff, Carol. "Sympathetic Magical Thinking: The Contagion and Similarity Heuristics," in Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds., *Heuristics and biases: The psychology of intuitive thought.* New York: Cambridge University Press, 2002, pp. 201–16.

Samuelson, William and Zeckhauser, Richard. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty*, March 1988, *1*(1), pp. 7–59.

Schelling, Thomas C. *Choice and consequence: Perspectives of an errant economist.* Cambridge, MA: Harvard University Press, 1984.

Schreiber, Charles A. and Kahneman, Daniel. "Determinants of the Remembered Utility of Aversive Sounds." *Journal of Experimental Psychology: General*, March 2000, *129*(1), pp. 27–42.

Shafir, Eldar and LeBoeuf, Robyn A. "Rationality." *Annual Review of Psychology*, February 2002, *53*(1), pp. 419–517.

Shiller, Robert J. *Irrational exuberance.* Princeton, NJ: Princeton University Press, 2000.

Shleifer, Andrei. *Inefficient markets: An introduction to behavioral finance.* New York: Oxford University Press, 2000.

Simon, Herbert A. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*, February 1955, *69*(1), pp. 99–118.

______. "Information Processing Models of Cognition." *Annual Review of Psychology*, February 1979, *30*, pp. 363–96.

Simon, Herbert A. and Chase, William G. "Skill in Chess." *American Scientist*, July 1973, *61*(4), pp. 394–403.

Sloman, Steven A. "Two Systems of Reasoning," in Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds., *Heuristics and biases: The psychology of intuitive thought.* New York: Cambridge University Press, 2002, pp. 379–96.

Slovic, Paul; Finucane, Melissa; Peters, Ellen and MacGregor, Donald G. "The Affect Heuristic," in Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds., *Heuristics and biases: The psychology of intuitive thought.* New York: Cambridge University Press, 2002, pp. 397–420.

Stanovich, Keith E. and West, Richard F. "Individual Differences in Reasoning: Implications for the Rationality Debate?" *Behavioral and Brain Sciences*, October 2000, *23*(5), pp. 645–65.

______. "Individual Differences in Reasoning: Implications for the Rationality Debate?" in Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds., *Heuristics and biases: The psychology of intuitive thought.* New York: Cambridge University Press, 2002, pp. 421–40.

Strack, Fritz; Martin, Leonard and Schwarz, Norbert. "Priming and Communication: Social Determinants of Information Use in Judgments of Life Satisfaction." *European Journal of Social Psychology*, October–November 1988, *18*(5), pp. 429–42.

Thaler, Richard H. "Toward a Positive Theory of Consumer Choice." *Journal of Economic*

Behavior and Organization, March 1980, *1*(1), pp. 36–90.

______. "Mental Accounting and Consumer Choice." *Marketing Science*, Summer 1985, *4*(3), pp. 199–214.

______. *Quasi rational economics*. New York: Russell Sage Foundation, 1991.

______. *The winner's curse: Paradoxes and anomalies of economic life*. New York: Free Press, 1992.

______. "Mental Accounting Matters." *Journal of Behavioral Decision Making*, July 1999, *12*(3), pp. 183–206.

______. "Toward a Positive Theory of Consumer Choice," in Daniel Kahneman and Amos Tversky, eds., *Choices, values, and frames*. New York: Cambridge University Press, 2000, pp. 268–87.

Tversky, Amos and Kahneman, Daniel. "Judgment under Uncertainty: Heuristics and Biases." *Science*, September 1974, *185*(4157), pp. 1124–31.

______. "The Framing of Decisions and the Psychology of Choice." *Science*, January 1981, *211*(4481), pp. 453–58.

______. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review*, October 1983, *90*(4), pp. 293–315.

______. "Rational Choice and the Framing of Decisions." *Journal of Business*, October 1986, *59*(4), pp. S251–78.

______. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics*, November 1991, *106*(4), pp. 1039–61.

______. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, October 1992, *5*(4), pp. 297–323.

Tversky, Amos and Redelmeier, Donald A. "On the Framing of Multiple Prospects." *Psychological Science*, May 1992, *3*(3), pp. 191–93.

Wilson, Timothy D. *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press, 2002.

Wilson, Timothy D. and Schooler, Jonathan W. "Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions." *Journal of Personality and Social Psychology*, February 1991, *60*(2), pp. 181–92.

Zajonc, Robert B. "Emotions," in Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, eds., *Handbook of social psychology*, 4th Ed., Vol. 1. New York: Oxford University Press, 1998, pp. 591–632.

Judgment under Uncertainty: Heuristics and Biases

Biases in judgments reveal some heuristics of thinking under uncertainty.

Amos Tversky and Daniel Kahneman

Many decisions are based on beliefs concerning the likelihood of uncertain events such as the outcome of an election, the guilt of a defendant, or the future value of the dollar. These beliefs are usually expressed in statements such as "I think that . . . ," "chances are . . . ," "it is unlikely that . . . ," and so forth. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective probabilities. What determines such beliefs? How do people assess the probability of an uncertain event or the value of an uncertain quantity? This article shows that people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.

The subjective assessment of probability resembles the subjective assessment of physical quantities such as distance or size. These judgments are all based on data of limited validity, which are processed according to heuristic rules. For example, the apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. This rule has some validity, because in any given scene the more distant objects are seen less sharply than nearer objects. However, the reliance on this rule leads to systematic errors in the estimation of distance. Specifically, distances are often overestimated when visibility is poor because the contours of objects are blurred. On the other hand, distances are often underestimated when visibility is good because the objects are seen sharply. Thus, the reliance on clarity as an indication of distance leads to common biases. Such biases are also found in the intuitive judgment of probability. This article describes three heuristics that are employed to assess probabilities and to predict values. Biases to which these heuristics lead are enumerated, and the applied and theoretical implications of these observations are discussed.

The authors are members of the department of psychology at the Hebrew University, Jerusalem, Israel.

Representativeness

Many of the probabilistic questions with which people are concerned belong to one of the following types: What is the probability that object A belongs to class B? What is the probability that event A originates from process B? What is the probability that process B will generate event A? In answering such questions, people typically rely on the representativeness heuristic, in which probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B. For example, when A is highly representative of B, the probability that A originates from B is judged to be high. On the other hand, if A is not similar to B, the probability that A originates from B is judged to be low.

For an illustration of judgment by representativeness, consider an individual who has been described by a former neighbor as follows: "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." How do people assess the probability that Steve is engaged in a particular occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely? In the representativeness heuristic, the probability that Steve is a librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian. Indeed, research with problems of this type has shown that people order the occupations by probability and by similarity in exactly the same way (*1*). This approach to the judgment of probability leads to serious errors, because similarity, or representativeness, is not influenced by several factors that should affect judgments of probability.

Insensitivity to prior probability of outcomes. One of the factors that have no effect on representativeness but should have a major effect on probability is the prior probability, or base-rate frequency, of the outcomes. In the case of Steve, for example, the fact that there are many more farmers than librarians in the population should enter into any reasonable estimate of the probability that Steve is a librarian rather than a farmer. Considerations of base-rate frequency, however, do not affect the similarity of Steve to the stereotypes of librarians and farmers. If people evaluate probability by representativeness, therefore, prior probabilities will be neglected. This hypothesis was tested in an experiment where prior probabilities were manipulated (*1*). Subjects were shown brief personality descriptions of several individuals, allegedly sampled at random from a group of 100 professionals—engineers and lawyers. The subjects were asked to assess, for each description, the probability that it belonged to an engineer rather than to a lawyer. In one experimental condition, subjects were told that the group from which the descriptions had been drawn consisted of 70 engineers and 30 lawyers. In another condition, subjects were told that the group consisted of 30 engineers and 70 lawyers. The odds that any particular description belongs to an engineer rather than to a lawyer should be higher in the first condition, where there is a majority of engineers, than in the second condition, where there is a majority of lawyers. Specifically, it can be shown by applying Bayes' rule that the ratio of these odds should be $(.7/.3)^2$, or 5.44, for each description. In a sharp violation of Bayes' rule, the subjects in the two conditions produced essentially the same probability judgments. Apparently, subjects evaluated the likelihood that a particular description belonged to an engineer rather than to a lawyer by the degree to which this description was representative of the two stereotypes, with little or no regard for the prior probabilities of the categories.

The subjects used prior probabilities correctly when they had no other information. In the absence of a personality sketch, they judged the probability that an unknown individual is an engineer to be .7 and .3, respectively, in the two base-rate conditions. However, prior probabilities were effectively ignored when a description was introduced, even when this description was totally uninformative. The responses to the following description illustrate this phenomenon:

> Dick is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

This description was intended to convey no information relevant to the question of whether Dick is an engineer or a lawyer. Consequently, the probability that Dick is an engineer should equal the proportion of engineers in the group, as if no description had been given. The subjects, however, judged the probability of Dick being an engineer to be .5 regardless of whether the stated proportion of engineers in the group was .7 or .3. Evidently, people respond differently when given no evidence and when given worthless evidence. When no specific evidence is given, prior probabilities are properly utilized; when worthless evidence is given, prior probabilities are ignored (*1*).

Insensitivity to sample size. To evaluate the probability of obtaining a particular result in a sample drawn from a specified population, people typically apply the representativeness heuristic. That is, they assess the likelihood of a sample result, for example, that the average height in a random sample of ten men will be 6 feet (180 centimeters), by the similarity of this result to the corresponding parameter (that is, to the average height in the population of men). The similarity of a sample statistic to a population parameter does not depend on the size of the sample. Consequently, if probabilities are assessed by representativeness, then the judged probability of a sample statistic will be essentially independent of sample size. Indeed, when subjects assessed the distributions of average height for samples of various sizes, they produced identical distributions. For example, the probability of obtaining an average height greater than 6 feet was assigned the same value for samples of 1000, 100, and 10 men (*2*). Moreover, subjects failed to appreciate the role of sample size even when it was emphasized in the formulation of the problem. Consider the following question:

> A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes lower.
>
> For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more such days?
>
> ► The larger hospital (21)
> ► The smaller hospital (21)
> ► About the same (that is, within 5 percent of each other) (53)

The values in parentheses are the number of undergraduate students who chose each answer.

Most subjects judged the probability of obtaining more than 60 percent boys to be the same in the small and in the large hospital, presumably because these events are described by the same statistic and are therefore equally representative of the general population. In contrast, sampling theory entails that the expected number of days on which more than 60 percent of the babies are boys is much greater in the small hospital than in the large one, because a large sample is less likely to stray from 50 percent. This fundamental notion of statistics is evidently not part of people's repertoire of intuitions.

A similar insensitivity to sample size has been reported in judgments of posterior probability, that is, of the probability that a sample has been drawn from one population rather than from another. Consider the following example:

> Imagine an urn filled with balls, of which ⅔ are of one color and ⅓ of another. One individual has drawn 5 balls from the urn, and found that 4 were red and 1 was white Another individual has drawn 20 balls and found that 12 were red and 8 were white. Which of the two individuals should feel more confident that the urn contains ⅔ red balls and ⅓ white balls, rather than the opposite? What odds should each individual give?

In this problem, the correct posterior odds are 8 to 1 for the 4 : 1 sample and 16 to 1 for the 12 : 8 sample, assuming equal prior probabilities. However, most people feel that the first sample provides much stronger evidence for the hypothesis that the urn is predominantly red, because the proportion of red balls is larger in the first than in the second sample. Here again, intuitive judgments are dominated by the sample proportion and are essentially unaffected by the size of the sample, which plays a crucial role in the determination of the actual posterior odds (*2*). In addition, intuitive estimates of posterior odds are far less extreme than the correct values. The underestimation of the impact of evidence has been observed repeatedly in problems of this type (*3*, *4*). It has been labeled "conservatism."

Misconceptions of chance. People expect that a sequence of events generated by a random process will represent the essential characteristics of that process even when the sequence is short. In considering tosses of a coin for heads or tails, for example, people regard the sequence H-T-H-T-T-H to be more likely than the sequence H-H-H-T-T-T, which does not appear random, and also more likely than the sequence H-H-H-H-T-H, which does not represent the fairness of the coin (*2*). Thus, people expect that the essential characteristics of the process will be represented, not only globally in the entire sequence, but also locally in each of its parts. A locally representative sequence, however, deviates systematically from chance expectation: it contains too many alternations and too few runs. Another consequence of the belief in local representativeness is the well-known gambler's fallacy. After observing a long run of red on the roulette wheel. for example, most people erroneously believe that black is now due, presumably because the occurrence of black will result in a more representative sequence than the occurrence of an additional red. Chance is commonly viewed as a self-correcting process in which a deviation in one direction induces a deviation in the opposite direction to restore the equilibrium. In fact, deviations are not "corrected" as a chance process unfolds, they are merely diluted.

Misconceptions of chance are not limited to naive subjects. A study of the statistical intuitions of experienced research psychologists (*5*) revealed a lingering belief in what may be called the "law of small numbers," according to which even small samples are highly representative of the populations from which they are drawn. The responses of these investigators reflected the expectation that a valid hypothesis about a population will be represented by a statistically significant result in a sample—with little regard for its size. As a consequence, the researchers put too much faith in the results of small samples and grossly overestimated the replicability of such results. In the actual conduct of research, this bias leads to the selection of samples of inadequate size and to overinterpretation of findings.

Insensitivity to predictability. People are sometimes called upon to make such numerical predictions as the future value of a stock, the demand for a commodity, or the outcome of a football game. Such predictions are often made by representativeness. For example, suppose one is given a description of a company and is asked to predict its future profit. If the description of the company is very favorable, a very high profit will appear most representative of that description; if the description is mediocre, a mediocre performance will appear most representative. The degree to which the description is favorable is unaffected by the reliability of that description or by the degree to which it permits accurate prediction. Hence, if people predict solely in terms of the favorableness of the description, their predictions will be insensitive to the reliability of the evidence and to the expected accuracy of the prediction.

This mode of judgment violates the normative statistical theory in which the extremeness and the range of predictions are controlled by considerations of predictability. When predictability is nil, the same prediction should be made in all cases. For example, if the descriptions of companies provide no information relevant to profit, then the same value (such as average profit) should be predicted for all companies. If predictability is perfect, of course, the values predicted will match the actual values and the range of predictions will equal the range of outcomes. In general, the higher the predictability, the wider the range of predicted values.

Several studies of numerical prediction have demonstrated that intuitive predictions violate this rule, and that subjects show little or no regard for considerations of predictability (*1*). In one of these studies, subjects were presented with several paragraphs, each describing the performance of a student teacher during a particular practice lesson. Some subjects were asked to *evaluate* the quality of the lesson described in the paragraph in percentile scores, relative to a specified population. Other subjects were asked to *predict*, also in percentile scores, the standing of each student teacher 5 years after the practice lesson. The judgments made under the two conditions were identical. That is, the prediction of a remote criterion (success of a teacher after 5 years) was identical to the evaluation of the information on which the prediction was based (the quality of the practice lesson). The students who made these predictions were undoubtedly aware of the limited predictability of teaching competence on the basis of a single trial lesson 5 years earlier; nevertheless, their predictions were as extreme as their evaluations.

The illusion of validity. As we have seen, people often predict by selecting the outcome (for example, an occupation) that is most representative of the input (for example, the description of a person). The confidence they have in their prediction depends primarily on the degree of representativeness (that is, on the quality of the match between the selected outcome and the input) with little or no regard for the factors that limit predictive accuracy. Thus, people express great confidence in the prediction that a person is a librarian when given a description of his personality which matches the stereotype of librarians, even if the description is scanty, unreliable, or outdated. The unwarranted confidence which is produced by a good fit between the predicted outcome and the input information may be called the illusion of validity. This illusion persists even when the judge is aware of the factors that limit the accuracy of his predictions. It is a common observation that psychologists who conduct selection interviews often experience considerable confidence in their predictions, even when they know of the vast literature that shows selection interviews to be highly fallible. The continued reliance on the clinical interview for selection, despite repeated demonstrations of its inadequacy, amply attests to the strength of this effect.

The internal consistency of a pattern of inputs is a major determinant of one's confidence in predictions based on these inputs. For example, people express more confidence in predicting the final grade-point average of a student whose first-year record consists entirely of B's than in predicting the grade-point average of a student whose first-year record includes many A's and C's. Highly consistent patterns are most often observed when the input variables are highly redundant or correlated. Hence, people tend to have great confidence in predictions based on redundant input variables. However, an elementary result in the statistics of correlation asserts that, given input variables of stated validity, a prediction based on several such inputs can achieve higher accuracy when they are independent of each other than when they are redundant or correlated. Thus, redundancy among inputs decreases accuracy even as it increases confidence, and people are often confident in predictions that are quite likely to be off the mark (*1*).

Misconceptions of regression. Suppose a large group of children has been examined on two equivalent versions of an aptitude test. If one selects ten children from among those who did best on one of the two versions, he will usually find their performance on the second version to be somewhat disappointing. Conversely, if one selects ten children from among those who did worst on one version, they will be found, on the average, to do somewhat better on the other version. More generally, consider two variables X and Y which have the same distribution. If one selects individuals whose average X score deviates from the mean of X by k units, then the average of their Y scores will usually deviate from the mean of Y by less than k units. These observations illustrate a general phenomenon known as regression toward the mean, which was first documented by Galton more than 100 years ago.

In the normal course of life, one encounters many instances of regression toward the mean, in the comparison of the height of fathers and sons, of the intelligence of husbands and wives, or of the performance of individuals on consecutive examinations. Nevertheless, people do not develop correct intuitions about this phenomenon. First, they do not expect regression in many contexts where it is bound to occur. Second, when they recognize the occurrence of regression, they often invent spurious causal explanations for it (*1*). We suggest that the phenomenon of regression remains elusive because it is incompatible with the belief that the predicted outcome should be maximally representative of the input, and, hence, that the value of the outcome variable should be as extreme as the value of the input variable.

The failure to recognize the import of regression can have pernicious consequences, as illustrated by the following observation (*1*). In a discussion of flight training, experienced instructors noted that praise for an exceptionally smooth landing is typically followed by a poorer landing on the next try, while harsh criticism after a rough landing is usually followed by an improvement on the next try. The instructors concluded that verbal rewards are detrimental to learning, while verbal punishments are beneficial, contrary to accepted psychological doctrine. This conclusion is unwarranted because of the presence of regression toward the mean. As in other cases of repeated examination, an improvement will usually follow a poor performance and a deterioration will usually follow an outstanding performance, even if the instructor does not respond to the trainee's achievement on the first attempt. Because the instructors had praised their trainees after good landings and admonished them after poor ones, they reached the erroneous and potentially harmful conclusion that punishment is more effective than reward.

Thus, the failure to understand the effect of regression leads one to overestimate the effectiveness of punishment and to underestimate the effectiveness of reward. In social interaction, as well as in training, rewards are typically administered when performance is good, and punishments are typically administered when performance is poor. By regression alone, therefore, behavior is most likely to improve after punishment and most likely to deteriorate after reward. Consequently, the human condition is such that, by chance alone, one is most often rewarded for punishing others and most often punished for rewarding them. People are generally not aware of this contingency. In fact, the elusive role of regression in determining the apparent consequences of reward and punishment seems to have escaped the notice of students of this area.

Availability

There are situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind. For example, one may assess the risk of heart attack among middle-aged people by recalling such occurrences among one's acquaintances. Similarly, one may evaluate the probability that a given business venture will fail by imagining various difficulties it could encounter. This judgmental heuristic is called availability. Availability is a useful clue for assessing frequency or probability, because instances of large classes are usually recalled better and faster than instances of less frequent classes. However, availability is affected by factors other than frequency and probability. Consequently, the reliance on availability leads to predictable biases, some of which are illustrated below.

Biases due to the retrievability of instances. When the size of a class is judged by the availability of its instances, a class whose instances are easily retrieved will appear more numerous than a class of equal frequency whose instances are less retrievable. In an elementary demonstration of this effect, subjects heard a list of well-known personalities of both sexes and were subsequently asked to judge whether the list contained more names of men than of women. Different lists were presented to different groups of subjects. In some of the lists the men were relatively more famous than the women, and in others the women were relatively more famous than the men. In each of the lists, the subjects erroneously judged that the class (sex) that had the more famous personalities was the more numerous (*6*).

In addition to familiarity, there are other factors, such as salience, which affect the retrievability of instances. For example, the impact of seeing a house burning on the subjective probability of such accidents is probably greater than the impact of reading about a fire in the local paper. Furthermore, recent occurrences are likely to be relatively more available than earlier occurrences. It is a common experience that the subjective probability of traffic accidents rises temporarily when one sees a car overturned by the side of the road.

Biases due to the effectiveness of a search set. Suppose one samples a word (of three letters or more) at random from an English text. Is it more likely that the word starts with r or that r is the third letter? People approach this problem by recalling words that begin with r (road) and words that have r in the third position (car) and assess the relative frequency by the ease with which words of the two types come to mind. Because it is much easier to search for words by their first letter than by their third letter, most people judge words that begin with a given consonant to be more numerous than words in which the same consonant appears in the third position. They do so even for consonants, such as r or k, that are more frequent in the third position than in the first (*6*).

Different tasks elicit different search sets. For example, suppose you are asked to rate the frequency with which abstract words (thought, love) and concrete words (door, water) appear in written English. A natural way to answer this question is to search for contexts in which the word could appear. It seems easier to think of contexts in which an abstract concept is mentioned (love in love stories) than to think of contexts in which a concrete word (such as door) is mentioned. If the frequency of words is judged by the availability of the contexts in which they appear, abstract words will be judged as relatively more numerous than concrete words. This bias has been observed in a recent study (*7*) which showed that the judged frequency of occurrence of abstract words was much higher than that of concrete words, equated in objective frequency. Abstract words were also judged to appear in a much greater variety of contexts than concrete words.

Biases of imaginability. Sometimes one has to assess the frequency of a class whose instances are not stored in memory but can be generated according to a given rule. In such situations, one typically generates several instances and evaluates frequency or probability by the ease with which the relevant instances can be constructed. However, the ease of constructing instances does not always reflect their actual frequency, and this mode of evaluation is prone to biases. To illustrate, consider a group of 10 people who form committees of k members, $2 \leqslant k \leqslant 8$. How many different committees of k members can be formed? The correct answer to this problem is given by the binomial coefficient $\binom{10}{k}$ which reaches a maximum of 252 for $k = 5$. Clearly, the number of committees of k members equals the number of committees of $(10 - k)$ members, because any committee of k

members defines a unique group of $(10 - k)$ nonmembers.

One way to answer this question without computation is to mentally construct committees of k members and to evaluate their number by the ease with which they come to mind. Committees of few members, say 2, are more available than committees of many members, say 8. The simplest scheme for the construction of committees is a partition of the group into disjoint sets. One readily sees that it is easy to construct five disjoint committees of 2 members, while it is impossible to generate even two disjoint committees of 8 members. Consequently, if frequency is assessed by imaginability, or by availability for construction, the small committees will appear more numerous than larger committees, in contrast to the correct bell-shaped function. Indeed, when naive subjects were asked to estimate the number of distinct committees of various sizes, their estimates were a decreasing monotonic function of committee size (*6*). For example, the median estimate of the number of committees of 2 members was 70, while the estimate for committees of 8 members was 20 (the correct answer is 45 in both cases).

Imaginability plays an important role in the evaluation of probabilities in real-life situations. The risk involved in an adventurous expedition, for example, is evaluated by imagining contingencies with which the expedition is not equipped to cope. If many such difficulties are vividly portrayed, the expedition can be made to appear exceedingly dangerous, although the ease with which disasters are imagined need not reflect their actual likelihood. Conversely, the risk involved in an undertaking may be grossly underestimated if some possible dangers are either difficult to conceive of, or simply do not come to mind.

Illusory correlation. Chapman and Chapman (*8*) have described an interesting bias in the judgment of the frequency with which two events co-occur. They presented naive judges with information concerning several hypothetical mental patients. The data for each patient consisted of a clinical diagnosis and a drawing of a person made by the patient. Later the judges estimated the frequency with which each diagnosis (such as paranoia or suspiciousness) had been accompanied by various features of the drawing (such as peculiar eyes). The subjects markedly overestimated the frequency of co-occurrence of natural associates, such as suspiciousness and peculiar eyes. This effect was labeled illusory correlation. In their erroneous judgments of the data to which they had been exposed, naive subjects "rediscovered" much of the common, but unfounded, clinical lore concerning the interpretation of the draw-a-person test. The illusory correlation effect was extremely resistant to contradictory data. It persisted even when the correlation between symptom and diagnosis was actually negative, and it prevented the judges from detecting relationships that were in fact present.

Availability provides a natural account for the illusory-correlation effect. The judgment of how frequently two events co-occur could be based on the strength of the associative bond between them. When the association is strong, one is likely to conclude that the events have been frequently paired. Consequently, strong associates will be judged to have occurred together frequently. According to this view, the illusory correlation between suspiciousness and peculiar drawing of the eyes, for example, is due to the fact that suspiciousness is more readily associated with the eyes than with any other part of the body.

Lifelong experience has taught us that, in general, instances of large classes are recalled better and faster than instances of less frequent classes; that likely occurrences are easier to imagine than unlikely ones; and that the associative connections between events are strengthened when the events frequently co-occur. As a result, man has at his disposal a procedure (the availability heuristic) for estimating the numerosity of a class, the likelihood of an event, or the frequency of co-occurrences, by the ease with which the relevant mental operations of retrieval, construction, or association can be performed. However, as the preceding examples have demonstrated, this valuable estimation procedure results in systematic errors.

Adjustment and Anchoring

In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting point, may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, adjustments are typically insufficient (*4*). That is, different starting points yield different estimates, which are biased toward the initial values. We call this phenomenon anchoring.

Insufficient adjustment. In a demonstration of the anchoring effect, subjects were asked to estimate various quantities, stated in percentages (for example, the percentage of African countries in the United Nations). For each quantity, a number between 0 and 100 was determined by spinning a wheel of fortune in the subjects' presence. The subjects were instructed to indicate first whether that number was higher or lower than the value of the quantity, and then to estimate the value of the quantity by moving upward or downward from the given number. Different groups were given different numbers for each quantity, and these arbitrary numbers had a marked effect on estimates. For example, the median estimates of the percentage of African countries in the United Nations were 25 and 45 for groups that received 10 and 65, respectively, as starting points. Payoffs for accuracy did not reduce the anchoring effect.

Anchoring occurs not only when the starting point is given to the subject, but also when the subject bases his estimate on the result of some incomplete computation. A study of intuitive numerical estimation illustrates this effect. Two groups of high school students estimated, within 5 seconds, a numerical expression that was written on the blackboard. One group estimated the product

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

while another group estimated the product

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

To rapidly answer such questions, people may perform a few steps of computation and estimate the product by extrapolation or adjustment. Because adjustments are typically insufficient, this procedure should lead to underestimation. Furthermore, because the result of the first few steps of multiplication (performed from left to right) is higher in the descending sequence than in the ascending sequence, the former expression should be judged larger than the latter. Both predictions were confirmed. The median estimate for the ascending sequence was 512, while the median estimate for the descending sequence was 2,250. The correct answer is 40,320.

Biases in the evaluation of conjunctive and disjunctive events. In a recent study by Bar-Hillel (*9*) subjects were given the opportunity to bet on one of two events. Three types of events were used: (i) simple events, such as drawing a red marble from a bag containing 50 percent red marbles and 50 percent white marbles; (ii) conjunctive events, such as drawing a red marble seven times in succession, with replacement, from a bag containing 90 percent red marbles and 10 percent white marbles; and (iii) disjunctive events, such as drawing a red marble at least once in seven successive tries, with replacement, from a bag containing 10 percent red marbles and 90 percent white marbles. In this problem, a significant majority of subjects preferred to bet on the conjunctive event (the probability of which is .48) rather than on the simple event (the probability of which is .50). Subjects also preferred to bet on the simple event rather than on the disjunctive event, which has a probability of .52. Thus, most subjects bet on the less likely event in both comparisons. This pattern of choices illustrates a general finding. Studies of choice among gambles and of judgments of probability indicate that people tend to overestimate the probability of conjunctive events (*10*) and to underestimate the probability of disjunctive events. These biases are readily explained as effects of anchoring. The stated probability of the elementary event (success at any one stage) provides a natural starting point for the estimation of the probabilities of both conjunctive and disjunctive events. Since adjustment from the starting point is typically insufficient, the final estimates remain too close to the probabilities of the elementary events in both cases. Note that the overall probability of a conjunctive event is lower than the probability of each elementary event, whereas the overall probability of a disjunctive event is higher than the probability of each elementary event. As a consequence of anchoring, the overall probability will be overestimated in conjunctive problems and underestimated in disjunctive problems.

Biases in the evaluation of compound events are particularly significant in the context of planning. The successful completion of an undertaking, such as the development of a new product, typically has a conjunctive character: for the undertaking to succeed, each of a series of events must occur. Even when each of these events is very likely, the overall probability of success can be quite low if the number of events is large. The general tendency to overestimate the probability of conjunctive events leads to unwarranted optimism in the evaluation of the likelihood that a plan will succeed or that a project will be completed on time. Conversely, disjunctive structures are typically encountered in the evaluation of risks. A complex system, such as a nuclear reactor or a human body, will malfunction if any of its essential components fails. Even when the likelihood of failure in each component is slight, the probability of an overall failure can be high if many components are involved. Because of anchoring, people will tend to underestimate the probabilities of failure in complex systems. Thus, the direction of the anchoring bias can sometimes be inferred from the structure of the event. The chain-like structure of conjunctions leads to overestimation, the funnel-like structure of disjunctions leads to underestimation.

Anchoring in the assessment of subjective probability distributions. In decision analysis, experts are often required to express their beliefs about a quantity, such as the value of the Dow-Jones average on a particular day, in the form of a probability distribution. Such a distribution is usually constructed by asking the person to select values of the quantity that correspond to specified percentiles of his subjective probability distribution. For example, the judge may be asked to select a number, X_{90}, such that his subjective probability that this number will be higher than the value of the Dow-Jones average is .90. That is, he should select the value X_{90} so that he is just willing to accept 9 to 1 odds that the Dow-Jones average will not exceed it. A subjective probability distribution for the value of the Dow-Jones average can be constructed from several such judgments corresponding to different percentiles.

By collecting subjective probability distributions for many different quantities, it is possible to test the judge for proper calibration. A judge is properly (or externally) calibrated in a set of problems if exactly Π percent of the true values of the assessed quantities falls below his stated values of X_{Π}. For example, the true values should fall below X_{01} for 1 percent of the quantities and above X_{99} for 1 percent of the quantities. Thus, the true values should fall in the confidence interval between X_{01} and X_{99} on 98 percent of the problems.

Several investigators (*11*) have obtained probability distributions for many quantities from a large number of judges. These distributions indicated large and systematic departures from proper calibration. In most studies, the actual values of the assessed quantities are either smaller than X_{01} or greater than X_{99} for about 30 percent of the problems. That is, the subjects state overly narrow confidence intervals which reflect more certainty than is justified by their knowledge about the assessed quantities. This bias is common to naive and to sophisticated subjects, and it is not eliminated by introducing proper scoring rules, which provide incentives for external calibration. This effect is attributable, in part at least, to anchoring.

To select X_{90} for the value of the Dow-Jones average, for example, it is natural to begin by thinking about one's best estimate of the Dow-Jones and to adjust this value upward. If this adjustment—like most others—is insufficient, then X_{90} will not be sufficiently extreme. A similar anchoring effect will occur in the selection of X_{10}, which is presumably obtained by adjusting one's best estimate downward. Consequently, the confidence interval between X_{10} and X_{90} will be too narrow, and the assessed probability distribution will be too tight. In support of this interpretation it can be shown that subjective probabilities are systematically altered by a procedure in which one's best estimate does not serve as an anchor.

Subjective probability distributions for a given quantity (the Dow-Jones average) can be obtained in two different ways: (i) by asking the subject to select values of the Dow-Jones that correspond to specified percentiles of his probability distribution and (ii) by asking the subject to assess the probabilities that the true value of the Dow-Jones will exceed some specified values. The two procedures are formally equivalent and should yield identical distributions. However, they suggest different modes of adjustment from different anchors. In procedure (i), the natural starting point is one's best estimate of the quantity. In procedure (ii), on the other hand, the subject may be anchored on the value stated in the question. Alternatively, he may be anchored on even odds, or 50-50 chances, which is a natural starting point in the estimation of likelihood. In either case, procedure (ii) should yield less extreme odds than procedure (i).

To contrast the two procedures, a set of 24 quantities (such as the air distance from New Delhi to Peking) was presented to a group of subjects who assessed either X_{10} or X_{90} for each problem. Another group of subjects received the median judgment of the first group for each of the 24 quantities. They were asked to assess the odds that each of the given values exceeded the true value of the relevant quantity. In the absence of any bias, the second group should retrieve the odds specified to the first group, that is, 9 : 1. However, if even odds or the stated value serve as anchors, the odds of the second group should be less extreme, that is, closer to 1 : 1. Indeed, the median odds stated by this group, across all problems, were 3 : 1. When the judgments of the two groups were tested for external calibration, it was found that subjects in the first group were too extreme, in accord with earlier studies. The events that they defined as having a probability of .10 actually obtained in 24 percent of the cases. In contrast, subjects in the second group were too conservative. Events to which they assigned an average probability of .34 actually obtained in 26 percent of the cases. These results illustrate the manner in which the degree of calibration depends on the procedure of elicitation.

Discussion

This article has been concerned with cognitive biases that stem from the reliance on judgmental heuristics. These biases are not attributable to motivational effects such as wishful thinking or the distortion of judgments by payoffs and penalties. Indeed, several of the severe errors of judgment reported earlier occurred despite the fact that subjects were encouraged to be accurate and were rewarded for the correct answers (*2, 6*).

The reliance on heuristics and the prevalence of biases are not restricted to laymen. Experienced researchers are also prone to the same biases—when they think intuitively. For example, the tendency to predict the outcome that best represents the data, with insufficient regard for prior probability, has been observed in the intuitive judgments of individuals who have had extensive training in statistics (*1, 5*). Although the statistically sophisticated avoid elementary errors, such as the gambler's fallacy, their intuitive judgments are liable to similar fallacies in more intricate and less transparent problems.

It is not surprising that useful heuristics such as representativeness and availability are retained, even though they occasionally lead to errors in prediction or estimation. What is perhaps surprising is the failure of people to infer from lifelong experience such fundamental statistical rules as regression toward the mean, or the effect of sample size on sampling variability. Although everyone is exposed, in the normal course of life, to numerous examples from which these rules could have been induced, very few people discover the principles of sampling and regression on their own. Statistical principles are not learned from everyday experience because the relevant instances are not coded appropriately. For example, people do not discover that successive lines in a text differ more in average word length than do successive pages, because they simply do not attend to the average word length of individual lines or pages. Thus, people do not learn the relation between sample size and sampling variability, although the data for such learning are abundant.

The lack of an appropriate code also explains why people usually do not detect the biases in their judgments of probability. A person could conceivably learn whether his judgments are externally calibrated by keeping a tally of the proportion of events that actually occur among those to which he assigns the same probability. However, it is not natural to group events by their judged probability. In the absence of such grouping it is impossible for an individual to discover, for example, that only 50 percent of the predictions to which he has assigned a probability of .9 or higher actually came true.

The empirical analysis of cognitive biases has implications for the theoretical and applied role of judged probabilities. Modern decision theory (*12, 13*) regards subjective probability as the quantified opinion of an idealized person. Specifically, the subjective probability of a given event is defined by the set of bets about this event that such a person is willing to accept. An internally consistent, or coherent, subjective probability measure can be derived for an individual if his choices among bets satisfy certain principles, that is, the axioms of the theory. The derived probability is subjective in the sense that different individuals are allowed to have different probabilities for the same event. The major contribution of this approach is that it provides a rigorous subjective interpretation of probability that is applicable to unique events and is embedded in a general theory of rational decision.

It should perhaps be noted that, while subjective probabilities can sometimes be inferred from preferences among bets, they are normally not formed in this fashion. A person bets on team A rather than on team B because he believes that team A is more likely to win; he does not infer this belief from his betting preferences. Thus, in reality, subjective probabilities determine preferences among bets and are not derived from them, as in the axiomatic theory of rational decision (*12*).

The inherently subjective nature of probability has led many students to the belief that coherence, or internal consistency, is the only valid criterion by which judged probabilities should be evaluated. From the standpoint of the formal theory of subjective probability, any set of internally consistent probability judgments is as good as any other. This criterion is not entirely satisfactory, because an internally consistent set of subjective probabilities can be incompatible with other beliefs held by the individual. Consider a person whose subjective probabilities for all possible outcomes of a coin-tossing game reflect the gambler's fallacy. That is, his estimate of the probability of tails on a particular toss increases with the number of consecutive heads that preceded that toss. The judgments of such a person could be internally consistent and therefore acceptable as adequate subjective probabilities according to the criterion of the formal theory. These probabilities, however, are incompatible with the generally held belief that a coin has no memory and is therefore incapable of generating sequential dependencies. For judged probabilities to be considered adequate, or rational, internal consistency is not enough. The judgments must be compatible with the entire web of beliefs held by the individual. Unfortunately, there can be no simple formal procedure for assessing the compatibility of a set of probability judgments with the judge's total system of beliefs. The rational judge will nevertheless strive for compatibility, even though internal consistency is more easily achieved and assessed. In particular, he will attempt to make his probability judgments compatible with his knowledge about the subject matter, the laws of probability, and his own judgmental heuristics and biases.

Summary

This article described three heuristics that are employed in making judgments under uncertainty: (i) representativeness, which is usually employed when people are asked to judge the probability that an object or event A belongs to class or process B; (ii) availability of instances or scenarios, which is often employed when people are asked to assess the frequency of a class or the plausibility of a particular development; and (iii) adjustment from an anchor, which is usually employed in numerical prediction when a relevant value is available. These heuristics are highly economical and usually effective, but they lead to systematic and predictable errors. A better understanding of these heuristics and of the biases to which they lead could improve judgments and decisions in situations of uncertainty.

References and Notes

1. D. Kahneman and A. Tversky, *Psychol. Rev.* **80**, 237 (1973).
2. ———, *Cognitive Psychol.* **3**, 430 (1972).
3. W. Edwards, in *Formal Representation of Human Judgment*, B. Kleinmuntz, Ed. (Wiley, New York, 1968), pp. 17–52.
4. P. Slovic and S. Lichtenstein, *Organ. Behav. Hum. Performance* **6**, 649 (1971).
5. A. Tversky and D. Kahneman, *Psychol. Bull.* **76**, 105 (1971).
6. ———, *Cognitive Psychol.* **5**, 207 (1973).
7. R. C. Galbraith and B. J. Underwood, *Mem. Cognition* **1**, 56 (1973).
8. L. J. Chapman and J. P. Chapman, *J. Abnorm. Psychol.* **73**, 193 (1967); *ibid.*, **74**, 271 (1969).
9. M. Bar-Hillel, *Organ. Behav. Hum. Performance* **9**, 396 (1973).
10. J. Cohen, E. I. Chesnick, D. Haran, *Br. J. Psychol.* **63**, 41 (1972).
11. M. Alpert and H. Raiffa, unpublished manuscript; C. A. S. von Holstein, *Acta Psychol.* **35**, 478 (1971); R. L. Winkler, *J. Am. Stat. Assoc.* **62**, 776 (1967).
12. L. J. Savage, *The Foundations of Statistics* (Wiley, New York, 1954).
13. B. De Finetti, in *International Encyclopedia of the Social Sciences*, D. E. Sills, Ed. (Macmillan, New York, 1968), vol. 12, pp. 496–504.

14. This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under contract N00014-73-C-0438 to the Oregon Research Institute, Eugene. Additional support for this research was provided by the Research and Development Authority of the Hebrew University, Jerusalem, Israel.

Rural Health Care in Mexico?

Present educational and administrative structures must be changed in order to improve health care in rural areas.

Luis Cañedo

The present health care structure in Mexico focuses attention on the urban population, leaving the rural communities practically unattended. There are two main factors contributing to this situation. One is the lack of coordination among the different institutions responsible for the health of the community and among the educational institutions. The other is the lack of information concerning the nature of the problems in rural areas. In an attempt to provide a solution to these problems, a program has been designed that takes into consideration the environmental conditions, malnutrition, poverty, and negative cultural factors that are responsible for the high incidences of certain diseases among rural populations. It is based on the development of a national information system for the collection and dissemination of information related to general, as well as rural, health care, that will provide the basis for a national health care system, and depends on the establishment of a training program for professionals in community medicine.

The continental and insular area of Mexico, including interior waters, is 2,022,058 square kilometers (*1, 2*). In 1970 the population of Mexico was 48,377,363, of which 24,055,305 persons (49.7 percent) were under 15 years of age. The Indian population made up 7.9 percent of the total (*2, 3*). As indicated in Table 1, 42.3 percent of the total population live in communities of less than 2,500 inhabitants, and in such communities public services as well as means of communication are very scarce or nonexistent. A large percentage (39.5 percent) of the economically active population is engaged in agriculture (*4*).

The country's population growth rate is high, 3.5 percent annually, and it seems to depend on income, being higher among the 50 percent of the population earning less than 675 pesos ($50) per family per month (*5*). The majority of this population lives in the rural areas. The most frequent causes of mortality in rural areas are malnutrition, infectious and parasitic diseases (*6, 7*), pregnancy complications, and accidents (*2*). In 1970 there were 34,107 doctors in Mexico (*2*). The ratio of inhabitants to doctors, which is 1423.7, is not a representative index of the actual distribution of resources because there is a great scarcity of health professionals in rural areas and a high concentration in urban areas (Fig. 1) (*7, 8*).

In order to improve health at a national level, this situation must be changed. The errors made in previous attempts to improve health care must be avoided, and use must be made of the available manpower and resources of modern science to produce feasible answers at the community level. Although the main objective of a specialist in community medicine is to control disease, such control cannot be achieved unless action is taken against the underlying causes of disease; it has already been observed that partial solutions are inefficient (*9*). As a background to this new program that has been designed to provide health care in rural communities, I shall first give a summary of the previous attempts that have been made to provide such care, describing the various medical institutions and other organizations that are responsible for the training of medical personnel and for constructing the facilities required for health care.

The author is an investigator in the department of molecular biology at the Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad Universitaria, México 20, D.F. This article is adapted from a paper presented at the meeting on Science and Man in the Americas, jointly organized by the Consejo Nacional de Ciencia y Tecnologia de México and the American Association for the Advancement of Science and held in Mexico City, 20 June to 4 July 1973.

References and Notes

[11] **The Assessment of Prior Distributions in Bayesian Analysis**
Robert L. Winkler
Journal of the American Statistical Association, Vol. 62, No. 319. (Sep., 1967), pp. 776-800.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28196709%2962%3A319%3C776%3ATAOPDI%3E2.0.CO%3B2-E

The Framing of Decisions and the Psychology of Choice

Amos Tversky and Daniel Kahneman

Explanations and predictions of people's choices, in everyday life as well as in the social sciences, are often founded on the assumption of human rationality. The definition of rationality has been much debated, but there is general agreement that rational choices should satisfy some elementary requirements of consistency and coherence. In this article we describe decision problems in which people systematically violate the requirements of consistency and coherence, and we trace these violations to the psychological principles that govern the perception of decision problems and the evaluation of options.

Summary. The psychological principles that govern the perception of decision problems and the evaluation of probabilities and outcomes produce predictable shifts of preference when the same problem is framed in different ways. Reversals of preference are demonstrated in choices regarding monetary outcomes, both hypothetical and real, and in questions pertaining to the loss of human lives. The effects of frames on preferences are compared to the effects of perspectives on perceptual appearance. The dependence of preferences on the formulation of decision problems is a significant concern for the theory of rational choice.

A decision problem is defined by the acts or options among which one must choose, the possible outcomes or consequences of these acts, and the contingencies or conditional probabilities that relate outcomes to acts. We use the term "decision frame" to refer to the decision-maker's conception of the acts, outcomes, and contingencies associated with a particular choice. The frame that a decision-maker adopts is controlled partly by the formulation of the problem and partly by the norms, habits, and personal characteristics of the decision-maker.

It is often possible to frame a given decision problem in more than one way. Alternative frames for a decision problem may be compared to alternative perspectives on a visual scene. Veridical perception requires that the perceived relative height of two neighboring mountains, say, should not reverse with changes of vantage point. Similarly, rational choice requires that the preference between options should not reverse with changes of frame. Because of imperfections of human perception and decision, however, changes of perspective often reverse the relative apparent size of objects and the relative desirability of options.

We have obtained systematic reversals of preference by variations in the framing of acts, contingencies, or outcomes. These effects have been observed in a variety of problems and in the choices of different groups of respondents. Here we present selected illustrations of preference reversals, with data obtained from students at Stanford University and at the University of British Columbia who answered brief questionnaires in a classroom setting. The total number of respondents for each problem is denoted by N, and the percentage who chose each option is indicated in brackets.

The effect of variations in framing is illustrated in problems 1 and 2.

Problem 1 [N = 152]: Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved. [72 percent]

If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [28 percent]

Which of the two programs would you favor?

The majority choice in this problem is risk averse: the prospect of certainly saving 200 lives is more attractive than a risky prospect of equal expected value, that is, a one-in-three chance of saving 600 lives.

A second group of respondents was given the cover story of problem 1 with a different formulation of the alternative programs, as follows:

Problem 2 [N = 155]:

If Program C is adopted 400 people will die. [22 percent]

If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. [78 percent]

Which of the two programs would you favor?

The majority choice in problem 2 is risk taking: the certain death of 400 people is less acceptable than the two-in-three chance that 600 will die. The preferences in problems 1 and 2 illustrate a common pattern: choices involving gains are often risk averse and choices involving losses are often risk taking. However, it is easy to see that the two problems are effectively identical. The only difference between them is that the outcomes are described in problem 1 by the number of lives saved and in problem 2 by the number of lives lost. The change is accompanied by a pronounced shift from risk aversion to risk taking. We have observed this reversal in several groups of respondents, including university faculty and physicians. Inconsistent responses to problems 1 and 2 arise from the conjunction of a framing effect with contradictory attitudes toward risks involving gains and losses. We turn now to an analysis of these attitudes.

The Evaluation of Prospects

The major theory of decision-making under risk is the expected utility model. This model is based on a set of axioms, for example, transitivity of preferences, which provide criteria for the rationality of choices. The choices of an individual who conforms to the axioms can be described in terms of the utilities of various outcomes for that individual. The utility of a risky prospect is equal to the expected utility of its outcomes, obtained by weighting the utility of each possible outcome by its probability. When faced with a choice, a rational decision-maker will prefer the prospect that offers the highest expected utility (*1*, *2*).

Dr. Tversky is a professor of psychology at Stanford University, Stanford, California 94305, and Dr. Kahneman is a professor of psychology at the University of British Columbia, Vancouver, Canada V6T 1W5.

0036-8075/81/0130-0453$01.50/0

As will be illustrated below, people exhibit patterns of preference which appear incompatible with expected utility theory. We have presented elsewhere (*3*) a descriptive model, called prospect theory, which modifies expected utility theory so as to accommodate these observations. We distinguish two phases in the choice process: an initial phase in which acts, outcomes, and contingencies are framed, and a subsequent phase of evaluation (*4*). For simplicity, we restrict the formal treatment of the theory to choices involving stated numerical probabilities and quantitative outcomes, such as money, time, or number of lives.

Consider a prospect that yields outcome x with probability p, outcome y with probability q, and the status quo with probability $1 - p - q$. According to prospect theory, there are values $v(.)$ associated with outcomes, and decision weights $\pi(.)$ associated with probabilities, such that the overall value of the prospect equals $\pi(p)\, v(x) + \pi(q)\, v(y)$. A slightly different equation should be applied if all outcomes of a prospect are on the same side of the zero point (*5*).

In prospect theory, outcomes are expressed as positive or negative deviations (gains or losses) from a neutral reference outcome, which is assigned a value of zero. Although subjective values differ among individuals and attributes, we propose that the value function is commonly S-shaped, concave above the reference point and convex below it, as illustrated in Fig. 1. For example, the difference in subjective value between gains of \$10 and \$20 is greater than the subjective difference between gains of \$110 and \$120. The same relation between value differences holds for the corresponding losses. Another property of the value function is that the response to losses is more extreme than the response to gains. The displeasure associated with losing a sum of money is generally greater than the pleasure associated with winning the same amount, as is reflected in people's reluctance to accept fair bets on a toss of a coin. Several studies of decision (*3, 6*) and judgment (*7*) have confirmed these properties of the value function (*8*).

The second major departure of prospect theory from the expected utility model involves the treatment of probabilities. In expected utility theory the utility of an uncertain outcome is weighted by its probability; in prospect theory the value of an uncertain outcome is multiplied by a decision weight $\pi(p)$, which is a monotonic function of p but is not a probability. The weighting function π

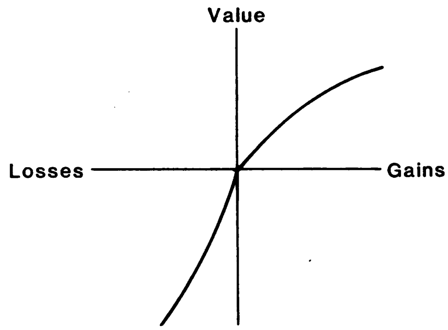


Fig. 1. A hypothetical value function.

has the following properties. First, impossible events are discarded, that is, $\pi(0) = 0$, and the scale is normalized so that $\pi(1) = 1$, but the function is not well behaved near the endpoints. Second, for low probabilities $\pi(p) > p$, but $\pi(p) + \pi(1 - p) \leq 1$. Thus low probabilities are overweighted, moderate and high probabilities are underweighted, and the latter effect is more pronounced than the former. Third, $\pi(pq)/\pi(p) < \pi(pqr)/\pi(pr)$ for all $0 < p, q, r \leq 1$. That is, for any fixed probability ratio q, the ratio of decision weights is closer to unity when the probabilities are low than when they are high, for example, $\pi(.1)/\pi(.2) > \pi(.4)/\pi(.8)$. A hypothetical weighting function which satisfies these properties is shown in Fig. 2. The major qualitative properties of decision weights can be extended to cases in which the probabilities of outcomes are subjectively assessed rather than explicitly given. In these situations, however, decision weights may also be affected by other characteristics of an event, such as ambiguity or vagueness (*9*).

Prospect theory, and the scales illustrated in Figs. 1 and 2, should be viewed as an approximate, incomplete, and simplified description of the evaluation of risky prospects. Although the properties of v and π summarize a common pattern of choice, they are not universal: the preferences of some individuals are not well described by an S-shaped value function and a consistent set of decision weights. The simultaneous measurement of values and decision weights involves serious experimental and statistical difficulties (*10*).

If π and v were linear throughout, the preference order between options would be independent of the framing of acts, outcomes, or contingencies. Because of the characteristic nonlinearities of π and v, however, different frames can lead to different choices. The following three sections describe reversals of preference caused by variations in the framing of acts, contingencies, and outcomes.

The Framing of Acts

Problem 3 [$N = 150$]: Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

Decision (i). Choose between:
- A. a sure gain of \$240 [84 percent]
- B. 25% chance to gain \$1000, and 75% chance to gain nothing [16 percent]

Decision (ii). Choose between:
- C. a sure loss of \$750 [13 percent]
- D. 75% chance to lose \$1000, and 25% chance to lose nothing [87 percent]

The majority choice in decision (i) is risk averse: a riskless prospect is preferred to a risky prospect of equal or greater expected value. In contrast, the majority choice in decision (ii) is risk taking: a risky prospect is preferred to a riskless prospect of equal expected value. This pattern of risk aversion in choices involving gains and risk seeking in choices involving losses is attributable to the properties of v and π. Because the value function is S-shaped, the value associated with a gain of \$240 is greater than 24 percent of the value associated with a gain of \$1000, and the (negative) value associated with a loss of \$750 is smaller than 75 percent of the value associated with a loss of \$1000. Thus the shape of the value function contributes to risk aversion in decision (i) and to risk seeking in decision (ii). Moreover, the underweighting of moderate and high probabilities contributes to the relative attractiveness of the sure gain in (i) and to the relative aversiveness of the sure loss in (ii). The same analysis applies to problems 1 and 2.

Because (i) and (ii) were presented together, the respondents had in effect to choose one prospect from the set: A and C, B and C, A and D, B and D. The most common pattern (A and D) was chosen by 73 percent of respondents, while the least popular pattern (B and C) was chosen by only 3 percent of respondents. However, the combination of B and C is definitely superior to the combination A and D, as is readily seen in problem 4.

Problem 4 [$N = 86$]. Choose between:
- A & D. 25% chance to win \$240, and 75% chance to lose \$760. [0 percent]
- B & C. 25% chance to win \$250, and 75% chance to lose \$750. [100 percent]

When the prospects were combined and the dominance of the second option became obvious, all respondents chose the superior option. The popularity of the inferior option in problem 3 implies that this problem was framed as a pair of separate choices. The respondents apparently failed to entertain the possibility that the conjunction of two seemingly reasonable choices could lead to an untenable result.

The violations of dominance observed in problem 3 do not disappear in the presence of monetary incentives. A different group of respondents who answered a modified version of problem 3, with real payoffs, produced a similar pattern of choices (*11*). Other authors have also reported that violations of the rules of rational choice, originally observed in hypothetical questions, were not eliminated by payoffs (*12*).

We suspect that many concurrent decisions in the real world are framed independently, and that the preference order would often be reversed if the decisions were combined. The respondents in problem 3 failed to combine options, although the integration was relatively simple and was encouraged by instructions (*13*). The complexity of practical problems of concurrent decisions, such as portfolio selection, would prevent people from integrating options without computational aids, even if they were inclined to do so.

The Framing of Contingencies

The following triple of problems illustrates the framing of contingencies. Each problem was presented to a different group of respondents. Each group was told that one participant in ten, preselected at random, would actually be playing for money. Chance events were realized, in the respondents' presence, by drawing a single ball from a bag containing a known proportion of balls of the winning color, and the winners were paid immediately.

Problem 5 [N = 77]: Which of the following options do you prefer?

A. a sure win of \$30 [78 percent]
B. 80% chance to win \$45 [22 percent]

Problem 6 [N = 85]: Consider the following two-stage game. In the first stage, there is a 75% chance to end the game without winning anything, and a 25% chance to move into the second stage. If you reach the second stage you have a choice between:

C. a sure win of \$30 [74 percent]
D. 80% chance to win \$45 [26 percent]

Your choice must be made before the game starts, i.e., before the outcome of the first stage is known. Please indicate the option you prefer.

Problem 7 [N = 81]: Which of the following options do you prefer?

E. 25% chance to win \$30 [42 percent]
F. 20% chance to win \$45 [58 percent]

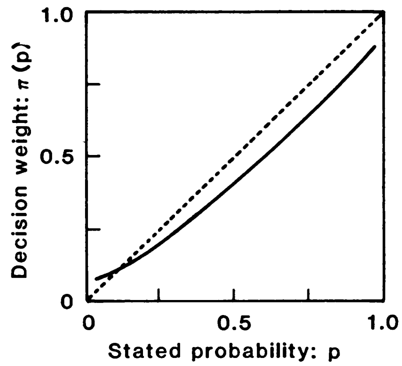


Fig. 2. A hypothetical weighting function.

Let us examine the structure of these problems. First, note that problems 6 and 7 are identical in terms of probabilities and outcomes, because prospect C offers a .25 chance to win \$30 and prospect D offers a probability of .25 × .80 = .20 to win \$45. Consistency therefore requires that the same choice be made in problems 6 and 7. Second, note that problem 6 differs from problem 5 only by the introduction of a preliminary stage. If the second stage of the game is reached, then problem 6 reduces to problem 5; if the game ends at the first stage, the decision does not affect the outcome. Hence there seems to be no reason to make a different choice in problems 5 and 6. By this logical analysis, problem 6 is equivalent to problem 7 on the one hand and problem 5 on the other. The participants, however, responded similarly to problems 5 and 6 but differently to problem 7. This pattern of responses exhibits two phenomena of choice: the certainty effect and the pseudocertainty effect.

The contrast between problems 5 and 7 illustrates a phenomenon discovered by Allais (*14*), which we have labeled the certainty effect: a reduction of the probability of an outcome by a constant factor has more impact when the outcome was initially certain than when it was merely probable. Prospect theory attributes this effect to the properties of π. It is easy to verify, by applying the equation of prospect theory to problems 5 and 7, that people for whom the value ratio $v(30)/v(45)$ lies between the weight ratios $\pi(.20)/\pi(.25)$ and $\pi(.80)/\pi(1.0)$ will prefer A to B and F to E, contrary to expected utility theory. Prospect theory does not predict a reversal of preference for every individual in problems 5 and 7. It only requires that an individual who has no preference between A and B prefer F to E. For group data, the theory predicts the observed directional shift of preference between the two problems.

The first stage of problem 6 yields the same outcome (no gain) for both acts. Consequently, we propose, people evaluate the options conditionally, as if the second stage had been reached. In this framing, of course, problem 6 reduces to problem 5. More generally, we suggest that a decision problem is evaluated conditionally when (i) there is a state in which all acts yield the same outcome, such as failing to reach the second stage of the game in problem 6, and (ii) the stated probabilities of other outcomes are conditional on the nonoccurrence of this state.

The striking discrepancy between the responses to problems 6 and 7, which are identical in outcomes and probabilities, could be described as a pseudocertainty effect. The prospect yielding \$30 is relatively more attractive in problem 6 than in problem 7, as if it had the advantage of certainty. The sense of certainty associated with option C is illusory, however, since the gain is in fact contingent on reaching the second stage of the game (*15*).

We have observed the certainty effect in several sets of problems, with outcomes ranging from vacation trips to the loss of human lives. In the negative domain, certainty exaggerates the aversiveness of losses that are certain relative to losses that are merely probable. In a question dealing with the response to an epidemic, for example, most respondents found "a sure loss of 75 lives" more aversive than "80% chance to lose 100 lives" but preferred "10% chance to lose 75 lives" over "8% chance to lose 100 lives," contrary to expected utility theory.

We also obtained the pseudocertainty effect in several studies where the description of the decision problems favored conditional evaluation. Pseudocertainty can be induced either by a sequential formulation, as in problem 6, or by the introduction of causal contingencies. In another version of the epidemic problem, for instance, respondents were told that risk to life existed only in the event (probability .10) that the disease was carried by a particular virus. Two alternative programs were said to yield "a sure loss of 75 lives" or "80% chance to lose 100 lives" if the critical virus was involved, and no loss of life in the event (probability .90) that the disease was carried by another virus. In effect, the respondents were asked to choose between 10 percent chance of losing 75 lives and 8 percent chance of losing 100 lives, but their preferences were the same as when the choice was ween a sure loss of 75 lives and 80 rcent chance of losing 100 lives. A nditional framing was evidently dopted in which the contingency of the noncritical virus was eliminated, giving rise to a pseudocertainty effect. The certainty effect reveals attitudes toward risk that are inconsistent with the axioms of rational choice, whereas the pseudocertainty effect violates the more fundamental requirement that preferences should be independent of problem description.

Many significant decisions concern actions that reduce or eliminate the probability of a hazard, at some cost. The shape of π in the range of low probabilities suggests that a protective action which reduces the probability of a harm from 1 percent to zero, say, will be valued more highly than an action that reduces the probability of the same harm from 2 percent to 1 percent. Indeed, probabilistic insurance, which reduces the probability of loss by half, is judged to be worth less than half the price of regular insurance that eliminates the risk altogether (*3*).

It is often possible to frame protective action in either conditional or unconditional form. For example, an insurance policy that covers fire but not flood could be evaluated either as full protection against the specific risk of fire or as a reduction in the overall probability of property loss. The preceding analysis suggests that insurance should appear more attractive when it is presented as the elimination of risk than when it is described as a reduction of risk. P. Slovic, B. Fischhoff, and S. Lichtenstein, in an unpublished study, found that a hypothetical vaccine which reduces the probability of contracting a disease from .20 to .10 is less attractive if it is described as effective in half the cases than if it is presented as fully effective against one of two (exclusive and equiprobable) virus strains that produce identical symptoms. In accord with the present analysis of pseudocertainty, the respondents valued full protection against an identified virus more than probabilistic protection against the disease.

The preceding discussion highlights the sharp contrast between lay responses to the reduction and the elimination of risk. Because no form of protective action can cover all risks to human welfare, all insurance is essentially probabilistic: it reduces but does not eliminate risk. The probabilistic nature of insurance is commonly masked by formulations that emphasize the completeness of protection against identified harms, but the sense of security that such formulations provide is an illusion of conditional framing. It appears that insurance is bought as protection against worry, not only against risk, and that worry can be manipulated by the labeling of outcomes and by the framing of contingencies. It is not easy to determine whether people value the elimination of risk too much or the reduction of risk too little. The contrasting attitudes to the two forms of protective action, however, are difficult to justify on normative grounds (*16*).

The Framing of Outcomes

Outcomes are commonly perceived as positive or negative in relation to a reference outcome that is judged neutral. Variations of the reference point can therefore determine whether a given outcome is evaluated as a gain or as a loss. Because the value function is generally concave for gains, convex for losses, and steeper for losses than for gains, shifts of reference can change the value difference between outcomes and thereby reverse the preference order between options (*6*). Problems 1 and 2 illustrated a preference reversal induced by a shift of reference that transformed gains into losses.

For another example, consider a person who has spent an afternoon at the race track, has already lost \$140, and is considering a \$10 bet on a 15:1 long shot in the last race. This decision can be framed in two ways, which correspond to two natural reference points. If the status quo is the reference point, the outcomes of the bet are framed as a gain of \$140 and a loss of \$10. On the other hand, it may be more natural to view the present state as a loss of \$140, for the betting day, and accordingly frame the last bet as a chance to return to the reference point or to increase the loss to \$150. Prospect theory implies that the latter frame will produce more risk seeking than the former. Hence, people who do not adjust their reference point as they lose are expected to take bets that they would normally find unacceptable. This analysis is supported by the observation that bets on long shots are most popular on the last race of the day (*17*).

Because the value function is steeper for losses than for gains, a difference between options will loom larger when it is framed as a disadvantage of one option rather than as an advantage of the other option. An interesting example of such an effect in a riskless context has been noted by Thaler (*18*). In a debate on a proposal to pass to the consumer some of the costs associated with the processing of credit-card purchases, representatives of the credit-card industry requested that the price difference be labeled a cash discount rather than a credit-card surcharge. The two labels induce different reference points by implicitly designating as normal reference the higher or the lower of the two prices. Because losses loom larger than gains, consumers are less willing to accept a surcharge than to forego a discount. A similar effect has been observed in experimental studies of insurance: the proportion of respondents who preferred a sure loss to a larger probable loss was significantly greater when the former was called an insurance premium (*19*, *20*).

These observations highlight the lability of reference outcomes, as well as their role in decision-making. In the examples discussed so far, the neutral reference point was identified by the labeling of outcomes. A diversity of factors determine the reference outcome in everyday life. The reference outcome is usually a state to which one has adapted; it is sometimes set by social norms and expectations; it sometimes corresponds to a level of aspiration, which may or may not be realistic.

We have dealt so far with elementary outcomes, such as gains or losses in a single attribute. In many situations, however, an action gives rise to a compound outcome, which joins a series of changes in a single attribute, such as a sequence of monetary gains and losses, or a set of concurrent changes in several attributes. To describe the framing and evaluation of compound outcomes, we use the notion of a psychological account, defined as an outcome frame which specifies (i) the set of elementary outcomes that are evaluated jointly and the manner in which they are combined and (ii) a reference outcome that is considered neutral or normal. In the account that is set up for the purchase of a car, for example, the cost of the purchase is not treated as a loss nor is the car viewed as a gift. Rather, the transaction as a whole is evaluated as positive, negative, or neutral, depending on such factors as the performance of the car and the price of similar cars in the market. A closely related treatment has been offered by Thaler (*18*).

We propose that people generally evaluate acts in terms of a minimal account, which includes only the direct consequences of the act. The minimal account associated with the decision to accept a gamble, for example, includes the money won or lost in that gamble and excludes other assets or the outcome of previous gambles. People commonly adopt minimal accounts because this mode of framing (i) simplifies evaluation and reduces cognitive strain, (ii) reflects the intuition that consequences should be causally linked to acts, and (iii) matches the properties of hedonic experience, which is more sensitive to desirable and undesirable changes than to steady states.

There are situations, however, in which the outcomes of an act affect the balance in an account that was previously set up by a related act. In these cases, the decision at hand may be evaluated in terms of a more inclusive account, as in the case of the bettor who views the last race in the context of earlier losses. More generally, a sunk-cost effect arises when a decision is referred to an existing account in which the current balance is negative. Because of the nonlinearities of the evaluation process, the minimal account and a more inclusive one often lead to different choices.

Problems 8 and 9 illustrate another class of situations in which an existing account affects a decision:

Problem 8 [N = 183]: Imagine that you have decided to see a play where admission is $10 per ticket. As you enter the theater you discover that you have lost a $10 bill.

Would you still pay $10 for a ticket for the play?

Yes [88 percent] No [12 percent]

Problem 9 [N = 200]: Imagine that you have decided to see a play and paid the admission price of $10 per ticket. As you enter the theater you discover that you have lost the ticket. The seat was not marked and the ticket cannot be recovered.

Would you pay $10 for another ticket?

Yes [46 percent] No [54 percent]

The marked difference between the responses to problems 8 and 9 is an effect of psychological accounting. We propose that the purchase of a new ticket in problem 9 is entered in the account that was set up by the purchase of the original ticket. In terms of this account, the expense required to see the show is $20, a cost which many of our respondents apparently found excessive. In problem 8, on the other hand, the loss of $10 is not linked specifically to the ticket purchase and its effect on the decision is accordingly slight.

The following problem, based on examples by Savage (*2*, p. 103) and Thaler (*18*), further illustrates the effect of embedding an option in different accounts. Two versions of this problem were presented to different groups of subjects. One group (N = 93) was given the values that appear in parentheses, and the other group (N = 88) the values shown in brackets.

Problem 10: Imagine that you are about to purchase a jacket for ($125) [$15], and a calculator for ($15) [$125]. The calculator salesman informs you that the calculator you wish to buy is on sale for ($10) [$120] at the other branch of the store, located 20 minutes drive away. Would you make the trip to the other store?

The response to the two versions of problem 10 were markedly different: 68 percent of the respondents were willing to make an extra trip to save $5 on a $15 calculator; only 29 percent were willing to exert the same effort when the price of the calculator was $125. Evidently the respondents do not frame problem 10 in the minimal account, which involves only a benefit of $5 and a cost of some inconvenience. Instead, they evaluate the potential saving in a more inclusive account, which includes the purchase of the calculator but not of the jacket. By the curvature of v, a discount of $5 has a greater impact when the price of the calculator is low than when it is high.

A closely related observation has been reported by Pratt, Wise, and Zeckhauser (*21*), who found that the variability of the prices at which a given product is sold by different stores is roughly proportional to the mean price of that product. The same pattern was observed for both frequently and infrequently purchased items. Overall, a ratio of 2:1 in the mean price of two products is associated with a ratio of 1.86:1 in the standard deviation of the respective quoted prices. If the effort that consumers exert to save each dollar on a purchase, for instance by a phone call, were independent of price, the dispersion of quoted prices should be about the same for all products. In contrast, the data of Pratt *et al.* (*21*) are consistent with the hypothesis that consumers hardly exert more effort to save $15 on a $150 purchase than to save $5 on a $50 purchase (*18*). Many readers will recognize the temporary devaluation of money which facilitates extra spending and reduces the significance of small discounts in the context of a large expenditure, such as buying a house or a car. This paradoxical variation in the value of money is incompatible with the standard analysis of consumer behavior.

Discussion

In this article we have presented a series of demonstrations in which seemingly inconsequential changes in the formulation of choice problems caused significant shifts of preference. The inconsistencies were traced to the interaction of two sets of factors: variations in the framing of acts, contingencies, and outcomes, and the characteristic nonlinearities of values and decision weights. The demonstrated effects are large and systematic, although by no means universal. They occur when the outcomes concern the loss of human lives as well as in choices about money; they are not restricted to hypothetical questions and are not eliminated by monetary incentives.

Earlier we compared the dependence of preferences on frames to the dependence of perceptual appearance on perspective. If while traveling in a mountain range you notice that the apparent relative height of mountain peaks varies with your vantage point, you will conclude that some impressions of relative height must be erroneous, even when you have no access to the correct answer. Similarly, one may discover that the relative attractiveness of options varies when the same decision problem is framed in different ways. Such a discovery will normally lead the decision-maker to reconsider the original preferences, even when there is no simple way to resolve the inconsistency. The susceptibility to perspective effects is of special concern in the domain of decision-making because of the absence of objective standards such as the true height of mountains.

The metaphor of changing perspective can be applied to other phenomena of choice, in addition to the framing effects with which we have been concerned here (*19*). The problem of self-control is naturally construed in these terms. The story of Ulysses' request to be bound to the mast of the ship in anticipation of the irresistible temptation of the Sirens' call is often used as a paradigm case (*22*). In this example of precommitment, an action taken in the present renders inoperative an anticipated future preference. An unusual feature of the problem of intertemporal conflict is that the agent who views a problem from a particular temporal perspective is also aware of the conflicting views that future perspectives will offer. In most other situations, decision-makers are not normally aware of the potential effects of different decision frames on their preferences.

The perspective metaphor highlights the following aspects of the psychology of choice. Individuals who face a decision problem and have a definite preference (i) might have a different preference in a different framing of the same problem, (ii) are normally unaware of alternative frames and of their potential effects on the relative attractiveness of options,

(iii) would wish their preferences to be independent of frame, but (iv) are often uncertain how to resolve detected inconsistencies (*23*). In some cases (such as problems 3 and 4 and perhaps problems 8 and 9) the advantage of one frame becomes evident once the competing frames are compared, but in other cases (problems 1 and 2 and problems 6 and 7) it is not obvious which preferences should be abandoned.

These observations do not imply that preference reversals, or other errors of choice or judgment (*24*), are necessarily irrational. Like other intellectual limitations, discussed by Simon (*25*) under the heading of "bounded rationality," the practice of acting on the most readily available frame can sometimes be justified by reference to the mental effort required to explore alternative frames and avoid potential inconsistencies. However, we propose that the details of the phenomena described in this article are better explained by prospect theory and by an analysis of framing than by ad hoc appeals to the notion of cost of thinking.

The present work has been concerned primarily with the descriptive question of how decisions are made, but the psychology of choice is also relevant to the normative question of how decisions ought to be made. In order to avoid the difficult problem of justifying values, the modern theory of rational choice has adopted the coherence of specific preferences as the sole criterion of rationality. This approach enjoins the decision-maker to resolve inconsistencies but offers no guidance on how to do so. It implicitly assumes that the decision-maker who carefully answers the question "What do I really want?" will eventually achieve coherent preferences. However, the susceptibility of preferences to variations of framing raises doubt about the feasibility and adequacy of the coherence criterion.

Consistency is only one aspect of the lay notion of rational behavior. As noted by March (*26*), the common conception of rationality also requires that preferences or utilities for particular outcomes should be predictive of the experiences of satisfaction or displeasure associated with their occurrence. Thus, a man could be judged irrational either because his preferences are contradictory or because his desires and aversions do not reflect his pleasures and pains. The predictive criterion of rationality can be applied to resolve inconsistent preferences and to improve the quality of decisions. A predictive orientation encourages the decision-maker to focus on future experience and to ask "What will I feel then?" rather than "What do I want now?" The former question, when answered with care, can be the more useful guide in difficult decisions. In particular, predictive considerations may be applied to select the decision frame that best represents the hedonic experience of outcomes.

Further complexities arise in the normative analysis because the framing of an action sometimes affects the actual experience of its outcomes. For example, framing outcomes in terms of overall wealth or welfare rather than in terms of specific gains and losses may attenuate one's emotional response to an occasional loss. Similarly, the experience of a change for the worse may vary if the change is framed as an uncompensated loss or as a cost incurred to achieve some benefit. The framing of acts and outcomes can also reflect the acceptance or rejection of responsibility for particular consequences, and the deliberate manipulation of framing is commonly used as an instrument of self-control (*22*). When framing influences the experience of consequences, the adoption of a decision frame is an ethically significant act.

References and Notes

1. J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, N.J., 1947); H. Raiffa, *Decision Analysis: Lectures on Choices Under Uncertainty* (Addison-Wesley, Reading, Mass., 1968); P. Fishburn, *Utility Theory for Decision Making* (Wiley, New York, 1970).
2. L. J. Savage, *The Foundations of Statistics* (Wiley, New York, 1954).
3. D. Kahneman and A. Tversky, *Econometrica* **47**, 263 (1979).
4. The framing phase includes various editing operations that are applied to simplify prospects, for example by combining events or outcomes or by discarding negligible components (*3*).
5. If $p + q = 1$ and either $x > y > 0$ or $x < y < 0$, the equation in the text is replaced by $v(y) + \pi(p)\ [v(x) - v(y)]$, so that decision weights are not applied to sure outcomes.
6. P. Fishburn and G. Kochenberger, *Decision Sci.* **10**, 503 (1979); D. J. Laughhunn, J. W. Payne, R. Crum, *Manage. Sci.*, in press; J. W. Payne, D. J. Laughhunn, R. Crum, *ibid.*, in press; S. A. Eraker and H. C. Sox, *Med. Decision Making*, in press. In the last study several hundred clinic patients made hypothetical choices between drug therapies for severe headaches, hypertension, and chest pain. Most patients were risk averse when the outcomes were described as positive (for example, reduced pain or increased life expectancy) and risk taking when the outcomes were described as negative (increased pain or reduced life expectancy). No significant differences were found between patients who actually suffered from the ailments described and patients who did not.
7. E. Galanter and P. Pliner, in *Sensation and Measurement*, H. R. Moskowitz *et al.*, Eds. (Reidel, Dordrecht, 1974), pp. 65–76.
8. The extension of the proposed value function to multiattribute options, with or without risk, deserves careful analysis. In particular, indifference curves between dimensions of loss may be concave upward, even when the value functions for the separate losses are both convex, because of marked subadditivity between dimensions.
9. D. Ellsberg, *Q. J. Econ.* **75**, 643 (1961); W. Fellner, *Probability and Profit—A Study of Economic Behavior Along Bayesian Lines* (Irwin, Homewood, Ill., 1965).
10. The scaling of v and π by pair comparisons requires a large number of observations. The procedure of pricing gambles is more convenient for scaling purposes, but it is subject to a severe anchoring bias: the ordering of gambles by their cash equivalents diverges systematically from the preference order observed in direct comparisons [S. Lichtenstein and P. Slovic, *J. Exp. Psychol.* **89**, 46 (1971)].
11. A new group of respondents ($N = 126$) was presented with a modified version of problem 3, in which the outcomes were reduced by a factor of 50. The participants were informed that the gambles would actually be played by tossing a pair of fair coins, that one participant in ten would be selected at random to play the gambles of his or her choice. To ensure a positive return for the entire set, a third decision, yielding only positive outcomes, was added. These payoff conditions did not alter the pattern of preferences observed in the hypothetical problem: 67 percent of respondents chose prospect A and 86 percent chose prospect D. The dominated combination of A and D was chosen by 60 percent of respondents, and only 6 percent favored the dominant combination of B and C.
12. S. Lichtenstein and P. Slovic, *J. Exp. Psychol.* **101**, 16 (1973); D. M. Grether and C. R. Plott, *Am. Econ. Rev.* **69**, 623 (1979); I. Lieblich and A. Lieblich, *Percept. Mot. Skills* **29**, 467 (1969); D. M. Grether, *Social Science Working Paper No. 245* (California Institute of Technology, Pasadena, 1979).
13. Other demonstrations of a reluctance to integrate concurrent options have been reported: P. Slovic and S. Lichtenstein, *J. Exp. Psychol.* **78**, 646 (1968); J. W. Payne and M. L. Braunstein, *ibid.* **87**, 13 (1971).
14. M. Allais, *Econometrica* **21**, 503 (1953); K. McCrimmon and S. Larsson, in *Expected Utility Hypotheses and the Allais Paradox*, M. Allais and O. Hagan, Eds. (Reidel, Dordrecht, 1979).
15. Another group of respondents ($N = 205$) was presented with all three problems, in different orders, without monetary payoffs. The joint frequency distribution of choices in problems 5, 6, and 7 was as follows: ACE, 22; ACF, 65; ADE, 4; ADF, 20; BCE, 7; BCF, 18; BDE, 17; BDF, 52. These data confirm in a within-subject design the analysis of conditional evaluation proposed in the text. More than 75 percent of respondents made compatible choices (AC or BD) in problems 5 and 6, and less than half made compatible choices in problems 6 and 7 (CE or DF) or 5 and 7 (AE or BF). The elimination of payoffs in these questions reduced risk aversion but did not substantially alter the effects of certainty and pseudocertainty.
16. For further discussion of rationality in protective action see H. Kunreuther, *Disaster Insurance Protection: Public Policy Lessons* (Wiley, New York, 1978).
17. W. H. McGlothlin, *Am. J. Psychol.* **69**, 604 (1956).
18. R. Thaler, *J. Econ. Behav. Organ.* **1**, 39 (1980).
19. B. Fischhoff, P. Slovic, S. Lichtenstein, in *Cognitive Processes in Choice and Decision Behavior*, T. Wallsten, Ed. (Erlbaum, Hillsdale, N.J., 1980).
20. J. C. Hershey and P. J. H. Schoemaker, *J. Risk Insur.*, in press.
21. J. Pratt, A. Wise, R. Zeckhauser, *Q. J. Econ.* **93**, 189 (1979).
22. R. H. Strotz, *Rev. Econ. Stud.* **23**, 165 (1955); G. Ainslie, *Psychol. Bull.* **82**, 463 (1975); J. Elster, *Ulysses and the Sirens: Studies in Rationality and Irrationality* (Cambridge Univ. Press, London, 1979); R. Thaler and H. M. Shifrin, *J. Polit. Econ.*, in press.
23. P. Slovic and A. Tversky, *Behav. Sci.* **19**, 368 (1974).
24. A. Tversky and D. Kahneman, *Science* **185**, 1124 (1974); P. Slovic, B. Fischhoff, S. Lichtenstein, *Annu. Rev. Psychol.* **28**, 1 (1977); R. Nisbett and L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Prentice-Hall, Englewood Cliffs, N.J., 1980); H. Einhorn and R. Hogarth, *Annu. Rev. Psychol.* **32**, 53 (1981).
25. H. A. Simon, *Q. J. Econ.* **69**, 99 (1955); *Psychol. Rev.* **63**, 129 (1956).
26. J. March, *Bell J. Econ.* **9**, 587 (1978).

27. This work was supported by the Office of Naval Research under contract N00014-79-C-0077 to Stanford University.

ECONOMETRICA

VOLUME 47 MARCH, 1979 NUMBER 2

PROSPECT THEORY: AN ANALYSIS OF DECISION UNDER RISK

BY DANIEL KAHNEMAN AND AMOS TVERSKY[1]

This paper presents a critique of expected utility theory as a descriptive model of decision making under risk, and develops an alternative model, called prospect theory. Choices among risky prospects exhibit several pervasive effects that are inconsistent with the basic tenets of utility theory. In particular, people underweight outcomes that are merely probable in comparison with outcomes that are obtained with certainty. This tendency, called the certainty effect, contributes to risk aversion in choices involving sure gains and to risk seeking in choices involving sure losses. In addition, people generally discard components that are shared by all prospects under consideration. This tendency, called the isolation effect, leads to inconsistent preferences when the same choice is presented in different forms. An alternative theory of choice is developed, in which value is assigned to gains and losses rather than to final assets and in which probabilities are replaced by decision weights. The value function is normally concave for gains, commonly convex for losses, and is generally steeper for losses than for gains. Decision weights are generally lower than the corresponding probabilities, except in the range of low probabilities. Overweighting of low probabilities may contribute to the attractiveness of both insurance and gambling.

1. INTRODUCTION

EXPECTED UTILITY THEORY has dominated the analysis of decision making under risk. It has been generally accepted as a normative model of rational choice [**24**], and widely applied as a descriptive model of economic behavior, e.g. [**15**, **4**]. Thus, it is assumed that all reasonable people would wish to obey the axioms of the theory [**47**, **36**], and that most people actually do, most of the time.

The present paper describes several classes of choice problems in which preferences systematically violate the axioms of expected utility theory. In the light of these observations we argue that utility theory, as it is commonly interpreted and applied, is not an adequate descriptive model and we propose an alternative account of choice under risk.

2. CRITIQUE

Decision making under risk can be viewed as a choice between prospects or gambles. A prospect $(x_1, p_1; \ldots; x_n, p_n)$ is a contract that yields outcome x_i with probability p_i, where $p_1+p_2+\ldots+p_n=1$. To simplify notation, we omit null outcomes and use (x, p) to denote the prospect $(x, p; 0, 1-p)$ that yields x with probability p and 0 with probability $1-p$. The (riskless) prospect that yields x with certainty is denoted by (x). The present discussion is restricted to prospects with so-called objective or standard probabilities.

The application of expected utility theory to choices between prospects is based on the following three tenets.

(i) Expectation: $U(x_1, p_1; \ldots; x_n, p_n) = p_1u(x_1)+\ldots+p_nu(x_n)$.

[1] This work was supported in part by grants from the Harry F. Guggenheim Foundation and from the Advanced Research Projects Agency of the Department of Defense and was monitored by Office of Naval Research under Contract N00014-78-C-0100 (ARPA Order No. 3469) under Subcontract 78-072-0722 from Decisions and Designs, Inc. to Perceptronics, Inc. We also thank the Center for Advanced Study in the Behavioral Sciences at Stanford for its support.

That is, the overall utility of a prospect, denoted by U, is the expected utility of its outcomes.

(ii) Asset Integration: $(x_1, p_1; \ldots; x_n, p_n)$ is acceptable at asset position w iff $U(w+x_1, p_1; \ldots; w+x_n, p_n) > u(w)$.

That is, a prospect is acceptable if the utility resulting from integrating the prospect with one's assets exceeds the utility of those assets alone. Thus, the domain of the utility function is final states (which include one's asset position) rather than gains or losses.

Although the domain of the utility function is not limited to any particular class of consequences, most applications of the theory have been concerned with monetary outcomes. Furthermore, most economic applications introduce the following additional assumption.

(iii) Risk Aversion: u is concave $(u'' < 0)$.

A person is risk averse if he prefers the certain prospect (x) to any risky prospect with expected value x. In expected utility theory, risk aversion is equivalent to the concavity of the utility function. The prevalence of risk aversion is perhaps the best known generalization regarding risky choices. It led the early decision theorists of the eighteenth century to propose that utility is a concave function of money, and this idea has been retained in modern treatments (Pratt [**33**], Arrow [**4**]).

In the following sections we demonstrate several phenomena which violate these tenets of expected utility theory. The demonstrations are based on the responses of students and university faculty to hypothetical choice problems. The respondents were presented with problems of the type illustrated below.

Which of the following would you prefer?

A: 50% chance to win 1,000,
50% chance to win nothing;

B: 450 for sure.

The outcomes refer to Israeli currency. To appreciate the significance of the amounts involved, note that the median net monthly income for a family is about 3,000 Israeli pounds. The respondents were asked to imagine that they were actually faced with the choice described in the problem, and to indicate the decision they would have made in such a case. The responses were anonymous, and the instructions specified that there was no 'correct' answer to such problems, and that the aim of the study was to find out how people choose among risky prospects. The problems were presented in questionnaire form, with at most a dozen problems per booklet. Several forms of each questionnaire were constructed so that subjects were exposed to the problems in different orders. In addition, two versions of each problem were used in which the left-right position of the prospects was reversed.

The problems described in this paper are selected illustrations of a series of effects. Every effect has been observed in several problems with different outcomes and probabilities. Some of the problems have also been presented to groups of students and faculty at the University of Stockholm and at the

University of Michigan. The pattern of results was essentially identical to the results obtained from Israeli subjects.

The reliance on hypothetical choices raises obvious questions regarding the validity of the method and the generalizability of the results. We are keenly aware of these problems. However, all other methods that have been used to test utility theory also suffer from severe drawbacks. Real choices can be investigated either in the field, by naturalistic or statistical observations of economic behavior, or in the laboratory. Field studies can only provide for rather crude tests of qualitative predictions, because probabilities and utilities cannot be adequately measured in such contexts. Laboratory experiments have been designed to obtain precise measures of utility and probability from actual choices, but these experimental studies typically involve contrived gambles for small stakes, and a large number of repetitions of very similar problems. These features of laboratory gambling complicate the interpretation of the results and restrict their generality.

By default, the method of hypothetical choices emerges as the simplest procedure by which a large number of theoretical questions can be investigated. The use of the method relies on the assumption that people often know how they would behave in actual situations of choice, and on the further assumption that the subjects have no special reason to disguise their true preferences. If people are reasonably accurate in predicting their choices, the presence of common and systematic violations of expected utility theory in hypothetical problems provides presumptive evidence against that theory.

Certainty, Probability, and Possibility

In expected utility theory, the utilities of outcomes are weighted by their probabilities. The present section describes a series of choice problems in which people's preferences systematically violate this principle. We first show that people overweight outcomes that are considered certain, relative to outcomes which are merely probable—a phenomenon which we label the *certainty effect.*

The best known counter-example to expected utility theory which exploits the certainty effect was introduced by the French economist Maurice Allais in 1953 [**2**]. Allais' example has been discussed from both normative and descriptive standpoints by many authors [**28**, **38**]. The following pair of choice problems is a variation of Allais' example, which differs from the original in that it refers to moderate rather than to extremely large gains. The number of respondents who answered each problem is denoted by N, and the percentage who choose each option is given in brackets.

PROBLEM 1: Choose between

A:	2,500 with probability	.33,	B:	2,400 with certainty.
	2,400 with probability	.66,		
	0 with probability	.01;		
$N = 72$	[18]			[82]*

PROBLEM 2: Choose between

C:	2,500 with probability	.33,	D:	2,400 with probability	.34,
	0 with probability	.67;		0 with probability	.66.
$N=72$	[83]*			[17]	

The data show that 82 per cent of the subjects chose B in Problem 1, and 83 per cent of the subjects chose C in Problem 2. Each of these preferences is significant at the .01 level, as denoted by the asterisk. Moreover, the analysis of individual patterns of choice indicates that a majority of respondents (61 per cent) made the modal choice in both problems. This pattern of preferences violates expected utility theory in the manner originally described by Allais. According to that theory, with $u(0)=0$, the first preference implies

$$u(2{,}400) > .33u(2{,}500)+.66u(2{,}400) \text{ or } .34u(2{,}400) > .33u(2{,}500)$$

while the second preference implies the reverse inequality. Note that Problem 2 is obtained from Problem 1 by eliminating a .66 chance of winning 2400 from both prospects under consideration. Evidently, this change produces a greater reduction in desirability when it alters the character of the prospect from a sure gain to a probable one, than when both the original and the reduced prospects are uncertain.

A simpler demonstration of the same phenomenon, involving only two-outcome gambles is given below. This example is also based on Allais [**2**].

PROBLEM 3:

A: (4,000,.80), or B: (3,000).

$N=95$ [20] [80]*

PROBLEM 4:

C: (4,000,.20), or D: (3,000,.25).

$N=95$ [65]* [35]

In this pair of problems as well as in all other problem-pairs in this section, over half the respondents violated expected utility theory. To show that the modal pattern of preferences in Problems 3 and 4 is not compatible with the theory, set $u(0)=0$, and recall that the choice of B implies $u(3{,}000)/u(4{,}000) > 4/5$, whereas the choice of C implies the reverse inequality. Note that the prospect $C=(4{,}000, .20)$ can be expressed as $(A, .25)$, while the prospect $D=(3{,}000, .25)$ can be rewritten as $(B,.25)$. The substitution axiom of utility theory asserts that if B is preferred to A, then any (probability) mixture (B, p) must be preferred to the mixture (A, p). Our subjects did not obey this axiom. Apparently, reducing the probability of winning from 1.0 to .25 has a greater effect than the reduction from

.8 to .2. The following pair of choice problems illustrates the certainty effect with non-monetary outcomes.

PROBLEM 5:

A: 50% chance to win a three-week tour of England, France, and Italy;

B: A one-week tour of England, with certainty.

$N = 72$ [22] [78]*

PROBLEM 6:

C: 5% chance to win a three-week tour of England, France, and Italy;

D: 10% chance to win a one-week tour of England.

$N = 72$ [67]* [33]

The certainty effect is not the only type of violation of the substitution axiom. Another situation in which this axiom fails is illustrated by the following problems.

PROBLEM 7:

A: (6,000, .45), B: (3,000, .90).

$N = 66$ [14] [86]*

PROBLEM 8:

C: (6,000, .001), D: (3,000, .002).

$N = 66$ [73]* [27]

Note that in Problem 7 the probabilities of winning are substantial (.90 and .45), and most people choose the prospect where winning is more probable. In Problem 8, there is a *possibility* of winning, although the probabilities of winning are minuscule (.002 and .001) in both prospects. In this situation where winning is possible but not probable, most people choose the prospect that offers the larger gain. Similar results have been reported by MacCrimmon and Larsson [**28**].

The above problems illustrate common attitudes toward risk or chance that cannot be captured by the expected utility model. The results suggest the following empirical generalization concerning the manner in which the substitution axiom is violated. If (y, pq) is equivalent to (x, p), then (y, pqr) is preferred to (x, pr), $0 < p, q, r < 1$. This property is incorporated into an alternative theory, developed in the second part of the paper.

The Reflection Effect

The previous section discussed preferences between positive prospects, i.e., prospects that involve no losses. What happens when the signs of the outcomes are reversed so that gains are replaced by losses? The left-hand column of Table I displays four of the choice problems that were discussed in the previous section, and the right-hand column displays choice problems in which the signs of the outcomes are reversed. We use $-x$ to denote the loss of x, and $>$ to denote the prevalent preference, i.e., the choice made by the majority of subjects.

TABLE I

PREFERENCES BETWEEN POSITIVE AND NEGATIVE PROSPECTS

	Positive prospects				Negative prospects		
Problem 3:	(4,000, .80)	<	(3,000).	Problem 3′:	(−4,000, .80)	>	(−3,000).
$N = 95$	[20]		[80]*	$N = 95$	[92]*		[8]
Problem 4:	(4,000, .20)	>	(3,000, .25).	Problem 4′:	(−4,000, .20)	<	(−3,000, .25).
$N = 95$	[65]*		[35]	$N = 95$	[42]		[58]
Problem 7:	(3,000, .90)	>	(6,000, .45).	Problem 7′:	(−3,000, .90)	<	(−6,000, .45).
$N = 66$	[86]*		[14]	$N = 66$	[8]		[92]*
Problem 8:	(3,000, .002)	<	(6,000, .001).	Problem 8′:	(−3,000, .002)	>	(−6,000, .001).
$N = 66$	[27]		[73]*	$N = 66$	[70]*		[30]

In each of the four problems in Table I the preference between negative prospects is the mirror image of the preference between positive prospects. Thus, the reflection of prospects around 0 reverses the preference order. We label this pattern the *reflection effect.*

Let us turn now to the implications of these data. First, note that the reflection effect implies that risk aversion in the positive domain is accompanied by risk seeking in the negative domain. In Problem 3′, for example, the majority of subjects were willing to accept a risk of .80 to lose 4,000, in preference to a sure loss of 3,000, although the gamble has a lower expected value. The occurrence of risk seeking in choices between negative prospects was noted early by Markowitz [**29**]. Williams [**48**] reported data where a translation of outcomes produces a dramatic shift from risk aversion to risk seeking. For example, his subjects were indifferent between (100, .65; −100, .35) and (0), indicating risk aversion. They were also indifferent between (−200, .80) and (−100), indicating risk seeking. A recent review by Fishburn and Kochenberger [**14**] documents the prevalence of risk seeking in choices between negative prospects.

Second, recall that the preferences between the positive prospects in Table I are inconsistent with expected utility theory. The preferences between the corresponding negative prospects also violate the expectation principle in the same manner. For example, Problems 3′ and 4′, like Problems 3 and 4, demonstrate that outcomes which are obtained with certainty are overweighted relative to uncertain outcomes. In the positive domain, the certainty effect contributes to a risk averse preference for a sure gain over a larger gain that is merely probable. In the negative domain, the same effect leads to a risk seeking preference for a loss that is merely probable over a smaller loss that is certain. The same psychological principle—the overweighting of certainty—favors risk aversion in the domain of gains and risk seeking in the domain of losses.

Third, the reflection effect eliminates aversion for uncertainty or variability as an explanation of the certainty effect. Consider, for example, the prevalent preferences for (3,000) over (4,000, .80) and for (4,000, .20) over (3,000, .25). To resolve this apparent inconsistency one could invoke the assumption that people prefer prospects that have high expected value and small variance (see, e.g., Allais [**2**]; Markowitz [**30**]; Tobin [**41**]). Since (3,000) has no variance while (4,000, .80) has large variance, the former prospect could be chosen despite its lower expected value. When the prospects are reduced, however, the difference in variance between (3,000, .25) and (4,000, .20) may be insufficient to overcome the difference in expected value. Because (−3,000) has both higher expected value and lower variance than (−4,000, .80), this account entails that the sure loss should be preferred, contrary to the data. Thus, our data are incompatible with the notion that certainty is generally desirable. Rather, it appears that certainty increases the aversiveness of losses as well as the desirability of gains.

Probabilistic Insurance

The prevalence of the purchase of insurance against both large and small losses has been regarded by many as strong evidence for the concavity of the utility function for money. Why otherwise would people spend so much money to purchase insurance policies at a price that exceeds the expected actuarial cost? However, an examination of the relative attractiveness of various forms of insurance does not support the notion that the utility function for money is concave everywhere. For example, people often prefer insurance programs that offer limited coverage with low or zero deductible over comparable policies that offer higher maximal coverage with higher deductibles—contrary to risk aversion (see, e.g., Fuchs [**16**]). Another type of insurance problem in which people's responses are inconsistent with the concavity hypothesis may be called probabilistic insurance. To illustrate this concept, consider the following problem, which was presented to 95 Stanford University students.

PROBLEM 9: Suppose you consider the possibility of insuring some property against damage, e.g., fire or theft. After examining the risks and the premium you find that you have no clear preference between the options of purchasing insurance or leaving the property uninsured.

It is then called to your attention that the insurance company offers a new program called *probabilistic insurance*. In this program you pay half of the regular premium. In case of damage, there is a 50 per cent chance that you pay the other half of the premium and the insurance company covers all the losses; and there is a 50 per cent chance that you get back your insurance payment and suffer all the losses. For example, if an accident occurs on an odd day of the month, you pay the other half of the regular premium and your losses are covered; but if the accident occurs on an even day of the month, your insurance payment is refunded and your losses are not covered.

Recall that the premium for full coverage is such that you find this insurance barely worth its cost.

Under these circumstances, would you purchase probabilistic insurance:

	Yes,	No.
$N=95$	[20]	[80]*

Although Problem 9 may appear contrived, it is worth noting that probabilistic insurance represents many forms of protective action where one pays a certain cost to reduce the probability of an undesirable event—without eliminating it altogether. The installation of a burglar alarm, the replacement of old tires, and the decision to stop smoking can all be viewed as probabilistic insurance.

The responses to Problem 9 and to several other variants of the same question indicate that probabilistic insurance is generally unattractive. Apparently, reducing the probability of a loss from p to $p/2$ is less valuable than reducing the probability of that loss from $p/2$ to 0.

In contrast to these data, expected utility theory (with a concave u) implies that probabilistic insurance is superior to regular insurance. That is, if at asset position w one is just willing to pay a premium y to insure against a probability p of losing x, then one should definitely be willing to pay a smaller premium ry to reduce the probability of losing x from p to $(1-r)p$, $0<r<1$. Formally, if one is indifferent between $(w-x, p; w, 1-p)$ and $(w-y)$, then one should prefer probabilistic insurance $(w-x, (1-r)p; w-y, rp; w-ry, 1-p)$ over regular insurance $(w-y)$.

To prove this proposition, we show that

$$pu(w-x)+(1-p)u(w)=u(w-y)$$

implies

$$(1-r)pu(w-x)+rpu(w-y)+(1-p)u(w-ry)>u(w-y).$$

Without loss of generality, we can set $u(w-x)=0$ and $u(w)=1$. Hence, $u(w-y)=1-p$, and we wish to show that

$$rp(1-p)+(1-p)u(w-ry)>1-p \qquad \text{or} \qquad u(w-ry)>1-rp$$

which holds if and only if u is concave.

This is a rather puzzling consequence of the risk aversion hypothesis of utility theory, because probabilistic insurance appears intuitively riskier than regular insurance, which entirely eliminates the element of risk. Evidently, the intuitive notion of risk is not adequately captured by the assumed concavity of the utility function for wealth.

The aversion for probabilistic insurance is particularly intriguing because all insurance is, in a sense, probabilistic. The most avid buyer of insurance remains vulnerable to many financial and other risks which his policies do not cover. There appears to be a significant difference between probabilistic insurance and what may be called contingent insurance, which provides the certainty of coverage for a specified type of risk. Compare, for example, probabilistic insurance against all forms of loss or damage to the contents of your home and contingent insurance that eliminates all risk of loss from theft, say, but does not cover other risks, e.g., fire. We conjecture that contingent insurance will be generally more attractive than probabilistic insurance when the probabilities of unprotected loss are equated. Thus, two prospects that are equivalent in probabilities and outcomes could have different values depending on their formulation. Several demonstrations of this general phenomenon are described in the next section.

The Isolation Effect

In order to simplify the choice between alternatives, people often disregard components that the alternatives share, and focus on the components that distinguish them (Tversky [**44**]). This approach to choice problems may produce inconsistent preferences, because a pair of prospects can be decomposed into common and distinctive components in more than one way, and different decompositions sometimes lead to different preferences. We refer to this phenomenon as the *isolation effect.*

PROBLEM 10: Consider the following two-stage game. In the first stage, there is a probability of .75 to end the game without winning anything, and a probability of .25 to move into the second stage. If you reach the second stage you have a choice between

(4,000, .80) and (3,000).

Your choice must be made before the game starts, i.e., before the outcome of the first stage is known.

Note that in this game, one has a choice between $.25\times.80=.20$ chance to win 4,000, and a $.25\times1.0=.25$ chance to win 3,000. Thus, in terms of final outcomes and probabilities one faces a choice between (4,000, .20) and (3,000, .25), as in Problem 4 above. However, the dominant preferences are different in the two problems. Of 141 subjects who answered Problem 10, 78 per cent chose the latter prospect, contrary to the modal preference in Problem 4. Evidently, people ignored the first stage of the game, whose outcomes are shared by both prospects, and considered Problem 10 as a choice between (3,000) and (4,000, .80), as in Problem 3 above.

The standard and the sequential formulations of Problem 4 are represented as decision trees in Figures 1 and 2, respectively. Following the usual convention, squares denote decision nodes and circles denote chance nodes. The essential difference between the two representations is in the location of the decision node. In the standard form (Figure 1), the decision maker faces a choice between two risky prospects, whereas in the sequential form (Figure 2) he faces a choice between a risky and a riskless prospect. This is accomplished by introducing a dependency between the prospects without changing either probabilities or

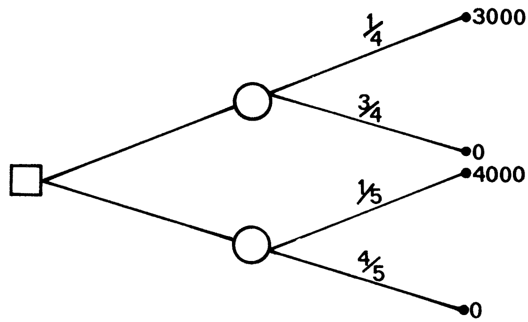


FIGURE 1.—The representation of Problem 4 as a decision tree (standard formulation).

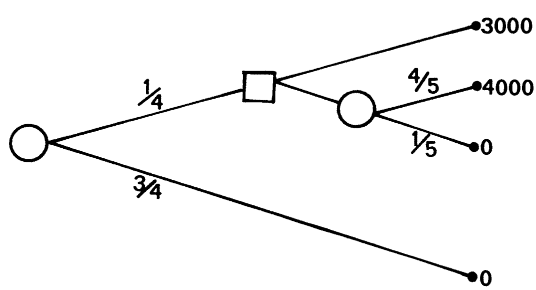


FIGURE 2.—The representation of Problem 10 as a decision tree (sequential formulation).

outcomes. Specifically, the event 'not winning 3,000' is included in the event 'not winning 4,000' in the sequential formulation, while the two events are independent in the standard formulation. Thus, the outcome of winning 3,000 has a certainty advantage in the sequential formulation, which it does not have in the standard formulation.

The reversal of preferences due to the dependency among events is particularly significant because it violates the basic supposition of a decision-theoretical analysis, that choices between prospects are determined solely by the probabilities of final states.

It is easy to think of decision problems that are most naturally represented in one of the forms above rather than in the other. For example, the choice between two different risky ventures is likely to be viewed in the standard form. On the other hand, the following problem is most likely to be represented in the sequential form. One may invest money in a venture with some probability of losing one's capital if the venture fails, and with a choice between a fixed agreed return and a percentage of earnings if it succeeds. The isolation effect implies that the contingent certainty of the fixed return enhances the attractiveness of this option, relative to a risky venture with the same probabilities and outcomes.

The preceding problem illustrated how preferences may be altered by different representations of probabilities. We now show how choices may be altered by varying the representation of outcomes.

Consider the following problems, which were presented to two different groups of subjects.

PROBLEM 11: In addition to whatever you own, you have been given 1,000. You are now asked to choose between

A: (1,000, .50), and B: (500).

$N=70$ [16] [84]*

PROBLEM 12: In addition to whatever you own, you have been given 2,000. You are now asked to choose between

C: (−1,000, .50), and D: (−500).

$N=68$ [69*] [31]

The majority of subjects chose *B* in the first problem and *C* in the second. These preferences conform to the reflection effect observed in Table I, which exhibits risk aversion for positive prospects and risk seeking for negative ones. Note, however, that when viewed in terms of final states, the two choice problems are identical. Specifically,

$$\mathrm{A} = (2{,}000, .50; 1{,}000, .50) = \mathrm{C}, \quad \text{and} \quad \mathrm{B} = (1{,}500) = \mathrm{D}.$$

In fact, Problem 12 is obtained from Problem 11 by adding 1,000 to the initial bonus, and subtracting 1,000 from all outcomes. Evidently, the subjects did not integrate the bonus with the prospects. The bonus did not enter into the comparison of prospects because it was common to both options in each problem.

The pattern of results observed in Problems 11 and 12 is clearly inconsistent with utility theory. In that theory, for example, the same utility is assigned to a wealth of $100, 000, regardless of whether it was reached from a prior wealth of $95,000 or $105,000. Consequently, the choice between a total wealth of $100,000 and even chances to own $95,000 or $105,000 should be independent of whether one currently owns the smaller or the larger of these two amounts. With the added assumption of risk aversion, the theory entails that the certainty of owning $100,000 should always be preferred to the gamble. However, the responses to Problem 12 and to several of the previous questions suggest that this pattern will be obtained if the individual owns the smaller amount, but not if he owns the larger amount.

The apparent neglect of a bonus that was common to both options in Problems 11 and 12 implies that the carriers of value or utility are changes of wealth, rather than final asset positions that include current wealth. This conclusion is the cornerstone of an alternative theory of risky choice, which is described in the following sections.

3. THEORY

The preceding discussion reviewed several empirical effects which appear to invalidate expected utility theory as a descriptive model. The remainder of the paper presents an alternative account of individual decision making under risk, called prospect theory. The theory is developed for simple prospects with monetary outcomes and stated probabilities, but it can be extended to more involved choices. Prospect theory distinguishes two phases in the choice process: an early phase of editing and a subsequent phase of evaluation. The editing phase consists of a preliminary analysis of the offered prospects, which often yields a simpler representation of these prospects. In the second phase, the edited prospects are evaluated and the prospect of highest value is chosen. We next outline the editing phase, and develop a formal model of the evaluation phase.

The function of the editing phase is to organize and reformulate the options so as to simplify subsequent evaluation and choice. Editing consists of the application of several operations that transform the outcomes and probabilities associated with the offered prospects. The major operations of the editing phase are described below.

Coding. The evidence discussed in the previous section shows that people normally perceive outcomes as gains and losses, rather than as final states of wealth or welfare. Gains and losses, of course, are defined relative to some neutral reference point. The reference point usually corresponds to the current asset position, in which case gains and losses coincide with the actual amounts that are received or paid. However, the location of the reference point, and the consequent coding of outcomes as gains or losses, can be affected by the formulation of the offered prospects, and by the expectations of the decision maker.

Combination. Prospects can sometimes be simplified by combining the probabilities associated with identical outcomes. For example, the prospect (200, .25; 200, .25) will be reduced to (200, .50). and evaluated in this form.

Segregation. Some prospects contain a riskless component that is segregated from the risky component in the editing phase. For example, the prospect (300, .80; 200, .20) is naturally decomposed into a sure gain of 200 and the risky prospect (100, .80). Similarly, the prospect (−400, .40; −100, .60) is readily seen to consist of a sure loss of 100 and of the prospect (−300, .40).

The preceding operations are applied to each prospect separately. The following operation is applied to a set of two or more prospects.

Cancellation. The essence of the isolation effects described earlier is the discarding of components that are shared by the offered prospects. Thus, our respondents apparently ignored the first stage of the sequential game presented in Problem 10, because this stage was common to both options, and they evaluated the prospects with respect to the results of the second stage (see Figure 2). Similarly, they neglected the common bonus that was added to the prospects in Problems 11 and 12. Another type of cancellation involves the discarding of common constituents, i.e., outcome-probability pairs. For example, the choice between (200, .20; 100, .50; −50, .30) and (200, .20; 150, .50; −100, .30) can be reduced by cancellation to a choice between (100, .50; −50, .30) and (150, .50; −100, .30).

Two additional operations that should be mentioned are simplification and the detection of dominance. The first refers to the simplification of prospects by rounding probabilities or outcomes. For example, the prospect (101, .49) is likely to be recoded as an even chance to win 100. A particularly important form of simplification involves the discarding of extremely unlikely outcomes. The second operation involves the scanning of offered prospects to detect dominated alternatives, which are rejected without further evaluation.

Because the editing operations facilitate the task of decision, it is assumed that they are performed whenever possible. However, some editing operations either permit or prevent the application of others. For example, (500, .20; 101, .49) will appear to dominate (500, .15; 99, .51) if the second constituents of both prospects are simplified to (100, .50). The final edited prospects could, therefore, depend on the sequence of editing operations, which is likely to vary with the structure of the offered set and with the format of the display. A detailed study of this problem is beyond the scope of the present treatment. In this paper we discuss choice problems where it is reasonable to assume either that the original formulation of the prospects leaves no room for further editing, or that the edited prospects can be specified without ambiguity.

Many anomalies of preference result from the editing of prospects. For example, the inconsistencies associated with the isolation effect result from the cancellation of common components. Some intransitivities of choice are explained by a simplification that eliminates small differences between prospects (see Tversky [**43**]). More generally, the preference order between prospects need not be invariant across contexts, because the same offered prospect could be edited in different ways depending on the context in which it appears.

Following the editing phase, the decision maker is assumed to evaluate each of the edited prospects, and to choose the prospect of highest value. The overall value of an edited prospect, denoted V, is expressed in terms of two scales, π and v.

The first scale, π, associates with each probability p a decision weight $\pi(p)$, which reflects the impact of p on the over-all value of the prospect. However, π is not a probability measure, and it will be shown later that $\pi(p)+\pi(1-p)$ is typically less than unity. The second scale, v, assigns to each outcome x a number $v(x)$, which reflects the subjective value of that outcome. Recall that outcomes are defined relative to a reference point, which serves as the zero point of the value scale. Hence, v measures the value of deviations from that reference point, i.e., gains and losses.

The present formulation is concerned with simple prospects of the form $(x, p; y, q)$, which have at most two non-zero outcomes. In such a prospect, one receives x with probability p, y with probability q, and nothing with probability $1-p-q$, where $p+q \leq 1$. An offered prospect is strictly positive if its outcomes are all positive, i.e., if $x, y > 0$ and $p+q=1$; it is strictly negative if its outcomes are all negative. A prospect is regular if it is neither strictly positive nor strictly negative.

The basic equation of the theory describes the manner in which π and v are combined to determine the over-all value of regular prospects.

If $(x, p; y, q)$ is a regular prospect (i.e., either $p+q<1$, or $x \geqslant 0 \geqslant y$, or $x \leqslant 0 \leqslant y$), then

$$V(x, p; y, q) = \pi(p)v(x) + \pi(q)v(y) \tag{1}$$

where $v(0)=0$, $\pi(0)=0$, and $\pi(1)=1$. As in utility theory, V is defined on prospects, while v is defined on outcomes. The two scales coincide for sure prospects, where $V(x, 1.0) = V(x) = v(x)$.

Equation (1) generalizes expected utility theory by relaxing the expectation principle. An axiomatic analysis of this representation is sketched in the Appendix, which describes conditions that ensure the existence of a unique π and a ratio-scale v satisfying equation (1).

The evaluation of strictly positive and strictly negative prospects follows a different rule. In the editing phase such prospects are segregated into two components: (i) the riskless component, i.e., the minimum gain or loss which is certain to be obtained or paid; (ii) the risky component, i.e., the additional gain or loss which is actually at stake. The evaluation of such prospects is described in the next equation.

If $p+q=1$ and either $x>y>0$ or $x<y<0$, then

$$V(x, p; y, q) = v(y) + \pi(p)[v(x) - v(y)]. \tag{2}$$

That is, the value of a strictly positive or strictly negative prospect equals the value of the riskless component plus the value-difference between the outcomes, multiplied by the weight associated with the more extreme outcome. For example, $V(400, .25; 100, .75) = v(100) + \pi(.25)[v(400) - v(100)]$. The essential feature of equation (2) is that a decision weight is applied to the value-difference $v(x) - v(y)$, which represents the risky component of the prospect, but not to $v(y)$, which represents the riskless component. Note that the right-hand side of equation (2) equals $\pi(p)v(x) + [1-\pi(p)]v(y)$. Hence, equation (2) reduces to equation (1) if $\pi(p) + \pi(1-p) = 1$. As will be shown later, this condition is not generally satisfied.

Many elements of the evaluation model have appeared in previous attempts to modify expected utility theory. Markowitz [**29**] was the first to propose that utility be defined on gains and losses rather than on final asset positions, an assumption which has been implicitly accepted in most experimental measurements of utility (see, e.g., [**7, 32**]). Markowitz also noted the presence of risk seeking in preferences among positive as well as among negative prospects, and he proposed a utility function which has convex and concave regions in both the positive and the negative domains. His treatment, however, retains the expectation principle; hence it cannot account for the many violations of this principle; see, e.g., Table I.

The replacement of probabilities by more general weights was proposed by Edwards [**9**], and this model was investigated in several empirical studies (e.g., [**3**, **42**]). Similar models were developed by Fellner [**12**], who introduced the concept of decision weight to explain aversion for ambiguity, and by van Dam [**46**] who attempted to scale decision weights. For other critical analyses of expected utility theory and alternative choice models, see Allais [**2**], Coombs [**6**], Fishburn [**13**], and Hansson [**22**].

The equations of prospect theory retain the general bilinear form that underlies expected utility theory. However, in order to accomodate the effects described in the first part of the paper, we are compelled to assume that values are attached to changes rather than to final states, and that decision weights do not coincide with stated probabilities. These departures from expected utility theory must lead to normatively unacceptable consequences, such as inconsistencies, intransitivities, and violations of dominance. Such anomalies of preference are normally corrected by the decision maker when he realizes that his preferences are inconsistent, intransitive, or inadmissible. In many situations, however, the decision maker does not have the opportunity to discover that his preferences could violate decision rules that he wishes to obey. In these circumstances the anomalies implied by prospect theory are expected to occur.

The Value Function

An essential feature of the present theory is that the carriers of value are changes in wealth or welfare, rather than final states. This assumption is compatible with basic principles of perception and judgment. Our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes. When we respond to attributes such as brightness, loudness, or temperature, the past and present context of experience defines an adaptation level, or reference point, and stimuli are perceived in relation to this reference point [**23**]. Thus, an object at a given temperature may be experienced as hot or cold to the touch depending on the temperature to which one has adapted. The same principle applies to non-sensory attributes such as health, prestige, and wealth. The same level of wealth, for example, may imply abject poverty for one person and great riches for another—depending on their current assets.

The emphasis on changes as the carriers of value should not be taken to imply that the value of a particular change is independent of initial position. Strictly speaking, value should be treated as a function in two arguments: the asset position that serves as reference point, and the magnitude of the change (positive or negative) from that reference point. An individual's attitude to money, say, could be described by a book, where each page presents the value function for changes at a particular asset position. Clearly, the value functions described on different pages are not identical: they are likely to become more linear with increases in assets. However, the preference order of prospects is not greatly altered by small or even moderate variations in asset position. The certainty equivalent of the prospect (1,000, .50), for example, lies between 300 and 400 for most people, in a wide range of asset positions. Consequently, the representation of value as a function in one argument generally provides a satisfactory approximation.

Many sensory and perceptual dimensions share the property that the psychological response is a concave function of the magnitude of physical change. For example, it is easier to discriminate between a change of 3° and a change of 6° in room temperature, than it is to discriminate between a change of 13° and a change of 16°. We propose that this principle applies in particular to the evaluation of monetary changes. Thus, the difference in value between a gain of 100 and a gain of 200 appears to be greater than the difference between a gain of 1,100 and a gain of 1,200. Similarly, the difference between a loss of 100 and a loss of 200 appears greater than the difference between a loss of 1,100 and a loss of 1,200, unless the larger loss is intolerable. Thus, we hypothesize that the value function for changes of wealth is normally concave above the reference point ($v''(x) < 0$, for $x > 0$) and often convex below it ($v''(x) > 0$, for $x < 0$). That is, the marginal value of both gains and losses generally decreases with their magnitude. Some support for this hypothesis has been reported by Galanter and Pliner [**17**], who scaled the perceived magnitude of monetary and non-monetary gains and losses.

The above hypothesis regarding the shape of the value function was based on responses to gains and losses in a riskless context. We propose that the value function which is derived from risky choices shares the same characteristics, as illustrated in the following problems.

PROBLEM 13:

	(6,000, .25),	or	(4,000, .25; 2,000, .25).
$N = 68$	[18]		[82]*

PROBLEM 13′:

	(−6,000, .25),	or	(−4,000, .25; −2,000, .25).
$N = 64$	[70]*		[30]

Applying equation 1 to the modal preference in these problems yields

$$\pi(.25)v(6{,}000) < \pi(.25)[v(4{,}000) + v(2{,}000)] \quad \text{and}$$
$$\pi(.25)v(-6{,}000) > \pi(.25)[v(-4{,}000) + v(-2{,}000)].$$

Hence, $v(6{,}000) < v(4{,}000) + v(2{,}000)$ and $v(-6{,}000) > v(-4{,}000) + v(-2{,}000)$. These preferences are in accord with the hypothesis that the value function is concave for gains and convex for losses.

Any discussion of the utility function for money must leave room for the effect of special circumstances on preferences. For example, the utility function of an individual who needs $60,000 to purchase a house may reveal an exceptionally steep rise near the critical value. Similarly, an individual's aversion to losses may increase sharply near the loss that would compel him to sell his house and move to a less desirable neighborhood. Hence, the derived value (utility) function of an individual does not always reflect "pure" attitudes to money, since it could be affected by additional consequences associated with specific amounts. Such perturbations can readily produce convex regions in the value function for gains and concave regions in the value function for losses. The latter case may be more common since large losses often necessitate changes in life style.

A salient characteristic of attitudes to changes in welfare is that losses loom larger than gains. The aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount [**17**]. Indeed, most people find symmetric bets of the form $(x, .50; -x, .50)$ distinctly unattractive. Moreover, the aversiveness of symmetric fair bets generally increases with the size of the stake. That is, if $x > y \geqslant 0$, then $(y, .50; -y, .50)$ is preferred to $(x, .50; -x, .50)$. According to equation (1), therefore,

$$v(y)+v(-y)>v(x)+v(-x) \qquad \text{and} \qquad v(-y)-v(-x)>v(x)-v(y).$$

Setting $y=0$ yields $v(x)<-v(-x)$, and letting y approach x yields $v'(x)< v'(-x)$, provided v', the derivative of v, exists. Thus, the value function for losses is steeper than the value function for gains.

In summary, we have proposed that the value function is (i) defined on deviations from the reference point; (ii) generally concave for gains and commonly convex for losses; (iii) steeper for losses than for gains. A value function which satisfies these properties is displayed in Figure 3. Note that the proposed S-shaped value function is steepest at the reference point, in marked contrast to the utility function postulated by Markowitz [**29**] which is relatively shallow in that region.

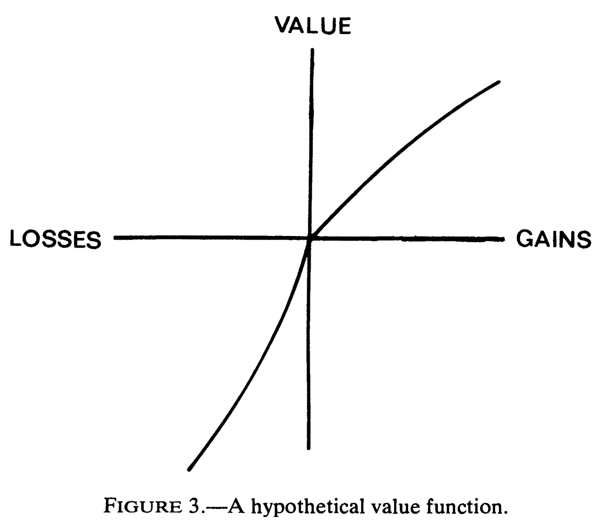


FIGURE 3.—A hypothetical value function.

Although the present theory can be applied to derive the value function from preferences between prospects, the actual scaling is considerably more complicated than in utility theory, because of the introduction of decision weights. For example, decision weights could produce risk aversion and risk seeking even with a linear value function. Nevertheless, it is of interest that the main properties ascribed to the value function have been observed in a detailed analysis of von Neumann–Morgenstern utility functions for changes of wealth (Fishburn and Kochenberger [**14**]). The functions had been obtained from thirty decision makers in various fields of business, in five independent studies [**5**, **18**, **19**, **21**, **40**]. Most utility functions for gains were concave, most functions for losses were convex, and only three individuals exhibited risk aversion for both gains and losses. With a single exception, utility functions were considerably steeper for losses than for gains.

The Weighting Function

In prospect theory, the value of each outcome is multiplied by a decision weight. Decision weights are inferred from choices between prospects much as subjective probabilities are inferred from preferences in the Ramsey-Savage approach. However, decision weights are not probabilities: they do not obey the probability axioms and they should not be interpreted as measures of degree or belief.

Consider a gamble in which one can win 1,000 or nothing, depending on the toss of a fair coin. For any reasonable person, the probability of winning is .50 in this situation. This can be verified in a variety of ways, e.g., by showing that the subject is indifferent between betting on heads or tails, or by his verbal report that he considers the two events equiprobable. As will be shown below, however, the decision weight $\pi(.50)$ which is derived from choices is likely to be smaller than .50. Decision weights measure the impact of events on the desirability of prospects, and not merely the perceived likelihood of these events. The two scales coincide (i.e., $\pi(p)=p$) if the expectation principle holds, but not otherwise.

The choice problems discussed in the present paper were formulated in terms of explicit numerical probabilities, and our analysis assumes that the respondents adopted the stated values of p. Furthermore, since the events were identified only by their stated probabilities, it is possible in this context to express decision weights as a function of stated probability. In general, however, the decision weight attached to an event could be influenced by other factors, e.g., ambiguity [**10**, **11**].

We turn now to discuss the salient properties of the weighting function π, which relates decision weights to stated probabilities. Naturally, π is an increasing function of p, with $\pi(0)=0$ and $\pi(1)=1$. That is, outcomes contingent on an impossible event are ignored, and the scale is normalized so that $\pi(p)$ is the ratio of the weight associated with the probability p to the weight associated with the certain event.

We first discuss some properties of the weighting function for small probabilities. The preferences in Problems 8 and 8′ suggest that for small values of p, π

is a subadditive function of p, i.e., $\pi(rp) > r\pi(p)$ for $0 < r < 1$. Recall that in Problem 8, (6,000, .001) is preferred to (3,000, .002). Hence

$$\frac{\pi(.001)}{\pi(.002)} > \frac{v(3{,}000)}{v(6{,}000)} > \frac{1}{2} \quad \text{by the concavity of } v.$$

The reflected preferences in Problem 8′ yield the same conclusion. The pattern of preferences in Problems 7 and 7′, however, suggests that subadditivity need not hold for large values of p.

Furthermore, we propose that very low probabilities are generally overweighted, that is, $\pi(p) > p$ for small p. Consider the following choice problems.

PROBLEM 14:

	(5,000, .001),	or	(5).
$N = 72$	[72]*		[28]

PROBLEM 14′:

	(−5,000, .001),	or	(−5).
$N = 72$	[17]		[83]*

Note that in Problem 14, people prefer what is in effect a lottery ticket over the expected value of that ticket. In Problem 14′, on the other hand, they prefer a small loss, which can be viewed as the payment of an insurance premium, over a small probability of a large loss. Similar observations have been reported by Markowitz [**29**]. In the present theory, the preference for the lottery in Problem 14 implies $\pi(.001)v(5{,}000) > v(5)$, hence $\pi(.001) > v(5)/v(5{,}000) > .001$, assuming the value function for gains is concave. The readiness to pay for insurance in Problem 14′ implies the same conclusion, assuming the value function for losses is convex.

It is important to distinguish overweighting, which refers to a property of decision weights, from the overestimation that is commonly found in the assessment of the probability of rare events. Note that the issue of overestimation does not arise in the present context, where the subject is assumed to adopt the stated value of p. In many real-life situations, overestimation and overweighting may both operate to increase the impact of rare events.

Although $\pi(p) > p$ for low probabilities, there is evidence to suggest that, for all $0 < p < 1$, $\pi(p) + \pi(1-p) < 1$. We label this property subcertainty. It is readily seen that the typical preferences in any version of Allias' example (see, e.g., Problems 1 and 2) imply subcertainty for the relevant value of p. Applying equation (1) to the prevalent preferences in Problems 1 and 2 yields, respectively,

$$v(2{,}400) > \pi(.66)v(2{,}400) + \pi(.33)v(2{,}500), \quad \text{i.e.,}$$

$$[1-\pi(.66]v(2{,}400) > \pi(.33)v(2{,}500) \quad \text{and}$$

$$\pi(.33)v(2{,}500) > \pi(.34)v(2{,}400); \quad \text{hence,}$$

$$1-\pi(.66) > \pi(.34), \quad \text{or} \quad \pi(.66)+\pi(.34) < 1.$$

Applying the same analysis to Allais' original example yields $\pi(.89)+\pi(.11)<1$, and some data reported by MacCrimmon and Larsson [**28**] imply subcertainty for additional values of p.

The slope of π in the interval (0, 1) can be viewed as a measure of the sensitivity of preferences to changes in probability. Subcertainty entails that π is regressive with respect to p, i.e., that preferences are generally less sensitive to variations of probability than the expectation principle would dictate. Thus, subcertainty captures an essential element of people's attitudes to uncertain events, namely that the sum of the weights associated with complementary events is typically less than the weight associated with the certain event.

Recall that the violations of the substitution axiom discussed earlier in this paper conform to the following rule: If (x, p) is equivalent to (y, pq) then (x, pr) is not preferred to (y, pqr), $0<p, q, r\leqslant 1$. By equation (1),

$$\pi(p)v(x)=\pi(pq)v(y) \quad \text{implies} \quad \pi(pr)v(x)\leqslant\pi(pqr)v(y); \quad \text{hence,}$$

$$\frac{\pi(pq)}{\pi(p)} \leqslant \frac{\pi(pqr)}{\pi(pr)}.$$

Thus, for a fixed ratio of probabilities, the ratio of the corresponding decision weights is closer to unity when the probabilities are low than when they are high. This property of π, called subproportionality, imposes considerable constraints on the shape of π: it holds if and only if $\log \pi$ is a convex function of $\log p$.

It is of interest to note that subproportionality together with the overweighting of small probabilities imply that π is subadditive over that range. Formally, it can be shown that if $\pi(p)>p$ and subproportionality holds, then $\pi(rp)>r\pi(p)$, $0<r<1$, provided π is monotone and continuous over (0, 1).

Figure 4 presents a hypothetical weighting function which satisfies overweighting and subadditivity for small values of p, as well as subcertainty and subproportionality. These properties entail that π is relatively shallow in the open interval and changes abruptly near the end-points where $\pi(0)=0$ and $\pi(1)=1$. The sharp drops or apparent discontinuities of π at the endpoints are consistent with the notion that there is a limit to how small a decision weight can be attached to an event, if it is given any weight at all. A similar quantum of doubt could impose an upper limit on any decision weight that is less than unity. This quantal effect may reflect the categorical distinction between certainty and uncertainty. On the other hand, the simplification of prospects in the editing phase can lead the individual to discard events of extremely low probability and to treat events of extremely high probability as if they were certain. Because people are limited in their ability to comprehend and evaluate extreme probabilities, highly unlikely events are either ignored or overweighted, and the difference between high probability and certainty is either neglected or exaggerated. Consequently, π is not well-behaved near the end-points.

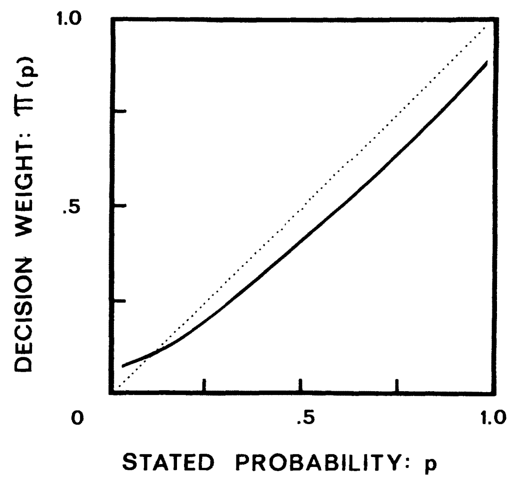


FIGURE 4.—A hypothetical weighting function.

The following example, due to Zeckhauser, illustrates the hypothesized nonlinearity of π. Suppose you are compelled to play Russian roulette, but are given the opportunity to purchase the removal of one bullet from the loaded gun. Would you pay as much to reduce the number of bullets from four to three as you would to reduce the number of bullets from one to zero? Most people feel that they would be willing to pay much more for a reduction of the probability of death from 1/6 to zero than for a reduction from 4/6 to 3/6. Economic considerations would lead one to pay more in the latter case, where the value of money is presumably reduced by the considerable probability that one will not live to enjoy it.

An obvious objection to the assumption that $\pi(p) \neq p$ involves comparisons between prospects of the form $(x, p; x, q)$ and $(x, p'; x, q')$, where $p+q=p'+q'<1$. Since any individual will surely be indifferent between the two prospects, it could be argued that this observation entails $\pi(p)+\pi(q)=\pi(p')+\pi(q')$, which in turn implies that π is the identity function. This argument is invalid in the present theory, which assumes that the probabilities of identical outcomes are combined in the editing of prospects. A more serious objection to the nonlinearity of π involves potential violations of dominance. Suppose $x>y>0$, $p>p'$, and $p+q=p'+q'<1$; hence, $(x, p; y, q)$ dominates $(x, p'; y, q')$. If preference obeys dominance, then

$$\pi(p)v(x)+\pi(q)v(y)>\pi(p')v(x)+\pi(q')v(y),$$

or

$$\frac{\pi(p)-\pi(p')}{\pi(q')-\pi(q)}>\frac{v(y)}{v(x)}.$$

Hence, as y approaches x, $\pi(p)-\pi(p')$ approaches $\pi(q')-\pi(q)$. Since $p-p'=q'-q$, π must be essentially linear, or else dominance must be violated.

Direct violations of dominance are prevented, in the present theory, by the assumption that dominated alternatives are detected and eliminated prior to the evaluation of prospects. However, the theory permits indirect violations of dominance, e.g., triples of prospects so that A is preferred to B, B is preferred to C, and C dominates A. For an example, see Raiffa [**34**, p. 75].

Finally, it should be noted that the present treatment concerns the simplest decision task in which a person chooses between two available prospects. We have not treated in detail the more complicated production task (e.g., bidding) where the decision maker generates an alternative that is equal in value to a given prospect. The asymmetry between the two options in this situation could introduce systematic biases. Indeed, Lichtenstein and Slovic [**27**] have constructed pairs of prospects A and B, such that people generally prefer A over B, but bid more for B than for A. This phenomenon has been confirmed in several studies, with both hypothetical and real gambles, e.g., Grether and Plott [**20**]. Thus, it cannot be generally assumed that the preference order of prospects can be recovered by a bidding procedure.

Because prospect theory has been proposed as a model of choice, the inconsistency of bids and choices implies that the measurement of values and decision weights should be based on choices between specified prospects rather than on bids or other production tasks. This restriction makes the assessment of v and π more difficult because production tasks are more convenient for scaling than pair comparisons.

4. DISCUSSION

In the final section we show how prospect theory accounts for observed attitudes toward risk, discuss alternative representations of choice problems induced by shifts of reference point, and sketch several extensions of the present treatment.

Risk Attitudes

The dominant pattern of preferences observed in Allais' example (Problems 1 and 2) follows from the present theory iff

$$\frac{\pi(.33)}{\pi(.34)}>\frac{v(2{,}400)}{v(2{,}500)}>\frac{\pi(.33)}{1-\pi(.66)}.$$

Hence, the violation of the independence axiom is attributed in this case to subcertainty, and more specifically to the inequality $\pi(.34) < 1 - \pi(.66)$. This analysis shows that an Allais-type violation will occur whenever the v-ratio of the two non-zero outcomes is bounded by the corresponding π-ratios.

Problems 3 through 8 share the same structure, hence it suffices to consider one pair, say Problems 7 and 8. The observed choices in these problems are implied by the theory iff

$$\frac{\pi(.001)}{\pi(.002)} > \frac{v(3{,}000)}{v(6{,}000)} > \frac{\pi(.45)}{\pi(.90)}.$$

The violation of the substitution axiom is attributed in this case to the subproportionality of π. Expected utility theory is violated in the above manner, therefore, whenever the v-ratio of the two outcomes is bounded by the respective π-ratios. The same analysis applies to other violations of the substitution axiom, both in the positive and in the negative domain.

We next prove that the preference for regular insurance over probabilistic insurance, observed in Problem 9, follows from prospect theory—provided the probability of loss is overweighted. That is, if $(-x, p)$ is indifferent to $(-y)$, then $(-y)$ is preferred to $(-x, p/2; -y, p/2; -y/2, 1-p)$. For simplicity, we define for $x \geqslant 0$, $f(x) = -v(-x)$. Since the value function for losses is convex, f is a concave function of x. Applying prospect theory, with the natural extension of equation 2, we wish to show that

$$\pi(p)f(x) = f(y) \qquad \text{implies}$$

$$\begin{aligned} f(y) &\leqslant f(y/2) + \pi(p/2)[f(y) - f(y/2)] + \pi(p/2)[f(x) - f(y/2)] \\ &= \pi(p/2)f(x) + \pi(p/2)f(y) + [1 - 2\pi(p/2)]f(y/2). \end{aligned}$$

Substituting for $f(x)$ and using the concavity of f, it suffices to show that

$$f(y) \leqslant \frac{\pi(p/2)}{\pi(p)} f(y) + \pi(p/2)f(y) + f(y)/2 - \pi(p/2)f(y)$$

or

$$\pi(p)/2 \leqslant \pi(p/2), \qquad \text{which follows from the subadditivity of } \pi.$$

According to the present theory, attitudes toward risk are determined jointly by v and π, and not solely by the utility function. It is therefore instructive to examine the conditions under which risk aversion or risk seeking are expected to occur. Consider the choice between the gamble (x, p) and its expected value (px). If $x > 0$, risk seeking is implied whenever $\pi(p) > v(px)/v(x)$, which is greater than p if the value function for gains is concave. Hence, overweighting $(\pi(p) > p)$ is necessary but not sufficient for risk seeking in the domain of gains. Precisely the same condition is necessary but not sufficient for risk aversion when $x < 0$. This analysis restricts risk seeking in the domain of gains and risk aversion in the domain of losses to small probabilities, where overweighting is expected to hold.

Indeed these are the typical conditions under which lottery tickets and insurance policies are sold. In prospect theory, the overweighting of small probabilities favors both gambling and insurance, while the S-shaped value function tends to inhibit both behaviors.

Although prospect theory predicts both insurance and gambling for small probabilities, we feel that the present analysis falls far short of a fully adequate account of these complex phenomena. Indeed, there is evidence from both experimental studies [**37**], survey research [**26**], and observations of economic behavior, e.g., service and medical insurance, that the purchase of insurance often extends to the medium range of probabilities, and that small probabilities of disaster are sometimes entirely ignored. Furthermore, the evidence suggests that minor changes in the formulation of the decision problem can have marked effects on the attractiveness of insurance [**37**]. A comprehensive theory of insurance behavior should consider, in addition to pure attitudes toward uncertainty and money, such factors as the value of security, social norms of prudence, the aversiveness of a large number of small payments spread over time, information and misinformation regarding probabilities and outcomes, and many others. Some effects of these variables could be described within the present framework, e.g., as changes of reference point, transformations of the value function, or manipulations of probabilities or decision weights. Other effects may require the introduction of variables or concepts which have not been considered in this treatment.

Shifts of Reference

So far in this paper, gains and losses were defined by the amounts of money that are obtained or paid when a prospect is played, and the reference point was taken to be the status quo, or one's current assets. Although this is probably true for most choice problems, there are situations in which gains and losses are coded relative to an expectation or aspiration level that differs from the status quo. For example, an unexpected tax withdrawal from a monthly pay check is experienced as a loss, not as a reduced gain. Similarly, an entrepreneur who is weathering a slump with greater success than his competitors may interpret a small loss as a gain, relative to the larger loss he had reason to expect.

The reference point in the preceding examples corresponded to an asset position that one had expected to attain. A discrepancy between the reference point and the current asset position may also arise because of recent changes in wealth to which one has not yet adapted [**29**]. Imagine a person who is involved in a business venture, has already lost 2,000 and is now facing a choice between a sure gain of 1,000 and an even chance to win 2,000 or nothing. If he has not yet adapted to his losses, he is likely to code the problem as a choice between $(-2{,}000, .50)$ and $(-1{,}000)$ rather than as a choice between $(2{,}000, .50)$ and $(1{,}000)$. As we have seen, the former representation induces more adventurous choices than the latter.

A change of reference point alters the preference order for prospects. In particular, the present theory implies that a negative translation of a choice problem, such as arises from incomplete adaptation to recent losses, increases risk seeking in some situations. Specifically, if a risky prospect $(x, p; -y, 1-p)$ is just acceptable, then $(x-z, p; -y-z, 1-p)$ is preferred over $(-z)$ for $x, y, z > 0$, with $x > z$.

To prove this proposition, note that

$$V(x, p; y, 1-p) = 0 \quad \text{iff} \quad \pi(p)v(x) = -\pi(1-p)v(-y).$$

Furthermore,

$$\begin{aligned}
&V(x-z, p; -y-z, 1-p) \\
&\quad = \pi(p)v(x-z) + \pi(1-p)v(-y-z) \\
&\quad > \pi(p)v(x) - \pi(p)v(z) + \pi(1-p)v(-y) \\
&\qquad + \pi(1-p)v(-z) \quad \text{by the properties of } v, \\
&\quad = -\pi(1-p)v(-y) - \pi(p)v(z) + \pi(1-p)v(-y) \\
&\qquad + \pi(1-p)v(-z) \quad \text{by substitution,} \\
&\quad = -\pi(p)v(z) + \pi(1-p)v(-z) \\
&\quad > v(-z)[\pi(p) + \pi(1-p)] \quad \text{since } v(-z) < -v(z), \\
&\quad > v(-z) \quad \text{by subcertainty.}
\end{aligned}$$

This analysis suggests that a person who has not made peace with his losses is likely to accept gambles that would be unacceptable to him otherwise. The well known observation [**31**] that the tendency to bet on long shots increases in the course of the betting day provides some support for the hypothesis that a failure to adapt to losses or to attain an expected gain induces risk seeking. For another example, consider an individual who expects to purchase insurance, perhaps because he has owned it in the past or because his friends do. This individual may code the decision to pay a premium y to protect against a loss x as a choice between $(-x+y, p; y, 1-p)$ and (0) rather than as a choice between $(-x, p)$ and $(-y)$. The preceding argument entails that insurance is likely to be more attractive in the former representation than in the latter.

Another important case of a shift of reference point arises when a person formulates his decision problem in terms of final assets, as advocated in decision analysis, rather than in terms of gains and losses, as people usually do. In this case, the reference point is set to zero on the scale of wealth and the value function is likely to be concave everywhere [**39**]. According to the present analysis, this formulation essentially eliminates risk seeking, except for gambling with low probabilities. The explicit formulation of decision problems in terms of final assets is perhaps the most effective procedure for eliminating risk seeking in the domain of losses.

Many economic decisions involve transactions in which one pays money in exchange for a desirable prospect. Current decision theories analyze such problems as comparisons between the status quo and an alternative state which includes the acquired prospect minus its cost. For example, the decision whether to pay 10 for the gamble (1,000, .01) is treated as a choice between (990, .01; −10, .99) and (0). In this analysis, readiness to purchase the positive prospect is equated to willingness to accept the corresponding mixed prospect.

The prevalent failure to integrate riskless and risky prospects, dramatized in the isolation effect, suggests that people are unlikely to perform the operation of subtracting the cost from the outcomes in deciding whether to buy a gamble. Instead, we suggest that people usually evaluate the gamble and its cost separately, and decide to purchase the gamble if the combined value is positive. Thus, the gamble (1,000, .01) will be purchased for a price of 10 if $\pi(.01)v(1{,}000)+v(-10)>0$.

If this hypothesis is correct, the decision to pay 10 for (1,000, .01), for example, is no longer equivalent to the decision to accept the gamble (990, .01; −10, .99). Furthermore, prospect theory implies that if one is indifferent between $(x(1-p), p; -px, 1-p)$ and (0) then one will not pay px to purchase the prospect (x, p). Thus, people are expected to exhibit more risk seeking in deciding whether to accept a fair gamble than in deciding whether to purchase a gamble for a fair price. The location of the reference point, and the manner in which choice problems are coded and edited emerge as critical factors in the analysis of decisions.

Extensions

In order to encompass a wider range of decision problems, prospect theory should be extended in several directions. Some generalizations are immediate; others require further development. The extension of equations (1) and (2) to prospects with any number of outcomes is straightforward. When the number of outcomes is large, however, additional editing operations may be invoked to simplify evaluation. The manner in which complex options, e.g., compound prospects, are reduced to simpler ones is yet to be investigated.

Although the present paper has been concerned mainly with monetary outcomes, the theory is readily applicable to choices involving other attributes, e.g., quality of life or the number of lives that could be lost or saved as a consequence of a policy decision. The main properties of the proposed value function for money should apply to other attributes as well. In particular, we expect outcomes to be coded as gains or losses relative to a neutral reference point, and losses to loom larger than gains.

The theory can also be extended to the typical situation of choice, where the probabilities of outcomes are not explicitly given. In such situations, decision weights must be attached to particular events rather than to stated probabilities, but they are expected to exhibit the essential properties that were ascribed to the weighting function. For example, if A and B are complementary events and neither is certain, $\pi(A)+\pi(B)$ should be less than unity—a natural analogue to subcertainty.

The decision weight associated with an event will depend primarily on the perceived likelihood of that event, which could be subject to major biases [**45**]. In addition, decision weights may be affected by other considerations, such as ambiguity or vagueness. Indeed, the work of Ellsberg [**10**] and Fellner [**12**] implies that vagueness reduces decision weights. Consequently, subcertainty should be more pronounced for vague than for clear probabilities.

The present analysis of preference between risky options has developed two themes. The first theme concerns editing operations that determine how prospects are perceived. The second theme involves the judgmental principles that govern the evaluation of gains and losses and the weighting of uncertain outcomes. Although both themes should be developed further, they appear to provide a useful framework for the descriptive analysis of choice under risk.

The University of British Columbia
and
Stanford University

Manuscript received November, 1977; final revision received March, 1978.

APPENDIX[2]

In this appendix we sketch an axiomatic analysis of prospect theory. Since a complete self-contained treatment is long and tedious, we merely outline the essential steps and exhibit the key ordinal properties needed to establish the bilinear representation of equation (1). Similar methods could be extended to axiomatize equation (2).

Consider the set of all regular prospects of the form $(x, p; y, q)$ with $p+q<1$. The extension to regular prospects with $p+q=1$ is straightforward. Let $\gtrsim$ denote the relation of preference between prospects that is assumed to be connected, symmetric and transitive, and let $\simeq$ denote the associated relation of indifference. Naturally, $(x, p; y, q)\simeq(y, q; x, p)$. We also assume, as is implicit in our notation, that $(x, p; 0, q)\simeq(x, p; 0, r)$, and $(x, p; y, 0)\simeq(x, p; z, 0)$. That is, the null outcome and the impossible event have the property of a multiplicative zero.

Note that the desired representation (equation (1)) is additive in the probability-outcome pairs. Hence, the theory of additive conjoint measurement can be applied to obtain a scale V which preserves the preference order, and interval scales f and g in two arguments such that

$$V(x, p; y, q)=f(x, p)+g(y, q).$$

The key axioms used to derive this representation are:

Independence: $(x, p; y, q)\gtrsim(x, p; y'q')$ iff $(x', p'; y, q)\gtrsim(x', p'; y', q')$.

Cancellation: If $(x, p; y'q')\gtrsim(x', p'; y, q)$ and $(x', p'; y'', q'')\gtrsim(x'', p''; y', q')$, then $(x, p; y'', q'')\gtrsim(x'', p''; y, q)$.

Solvability: If $(x, p; y, q)\gtrsim(z, r)\gtrsim(x, p; y' q')$ for some outcome z and probability r, then there exist y'', q'' such that

$$(x, p; y''q'')\simeq(z, r).$$

It has been shown that these conditions are sufficient to construct the desired additive representation, provided the preference order is Archimedean [**8, 25**]. Furthermore, since $(x, p; y, q)\simeq(y, q; x, p)$, $f(x, p)+g(y, q)=f(y, q)+g(x, p)$, and letting $q=0$ yields $f=g$.

Next, consider the set of all prospects of the form (x, p) with a single non-zero outcome. In this case, the bilinear model reduces to $V(x, p)=\pi(p)v(x)$. This is the multiplicative model, investigated in [**35**] and [**25**]. To construct the multiplicative representation we assume that the ordering of the probability-outcome pairs satisfies independence, cancellation, solvability, and the Archimedean axiom. In addition, we assume sign dependence [**25**] to ensure the proper multiplication of signs. It should be noted that the solvability axiom used in [**35**] and [**25**] must be weakened because the probability factor permits only bounded solvability.

[2] We are indebted to David H. Krantz for his help in the formulation of this section.

Combining the additive and the multiplicative representations yields

$$V(x, p; y, q) = f[\pi(p)v(x)] + f[\pi(q)v(y)].$$

Finally, we impose a new distributivity axiom:

$$(x, p; y, p) \simeq (z, p) \quad \text{iff} \quad (x, q; y, q) \simeq (z, q).$$

Applying this axiom to the above representation, we obtain

$$f[\pi(p)v(x)] + f[\pi(p)v(y)] = f[\pi(p)v(z)]$$

implies

$$f[\pi(q)v(x)] + f[\pi(q)v(y)] = f[\pi(q)v(z)].$$

Assuming, with no loss of generality, that $\pi(q) < \pi(p)$, and letting $\alpha = \pi(p)v(x)$, $\beta = \pi(p)v(y)$, $\gamma = \pi(p)v(z)$, and $\theta = \pi(q)/\pi(p)$, yields $f(\alpha) + f(\beta) = f(\gamma)$ implies $f(\theta\alpha) + f(\theta\beta) = f(\theta\gamma)$ for all $0 < \theta < 1$.

Because f is strictly monotonic we can set $\gamma = f^{-1}[f(\alpha) + f(\beta)]$. Hence, $\theta\gamma = \theta f^{-1}[f(\alpha) + f(\beta)] = f^{-1}[f(\theta\alpha) + f(\theta\beta)]$.

The solution to this functional equation is $f(\alpha) = k\alpha^c$ [1]. Hence, $V(x, p; y, q) = k[\pi(p)v(x)]^c + k[\pi(q)v(y)]^c$, for some $k, c > 0$. The desired bilinear form is obtained by redefining the scales π, v, and V so as to absorb the constants k and c.

REFERENCES

[1] Aczél, J.: *Lectures on Functional Equations and Their Applications.* New York: Academic Press, 1966.

[2] Allais, M.: "Le Comportement de l'Homme Rationnel devant le Risque, Critique des Postulats et Axiomes de l'Ecole Americaine," *Econometrica*, 21 (1953), 503–546.

[3] Anderson, N. H., and J. C. Shanteau: "Information Integration in Risky Decision Making," *Journal of Experimental Psychology*, 84 (1970), 441–451.

[4] Arrow, K. J.: *Essays in the Theory of Risk-Bearing.* Chicago: Markham, 1971.

[5] Barnes, J. D., and J. E. Reinmuth: "Comparing Imputed and Actual Utility Functions in a Competitive Bidding Setting," *Decision Sciences*, 7 (1976), 801–812.

[6] Coombs, C. H.: "Portfolio Theory and the Measurement of Risk," in *Human Judgment and Decision Processes*, ed. by M. F. Kaplan and S. Schwartz. New York: Academic Press, 1975, pp. 63–85.

[7] Davidson, D., P. Suppes, and S. Siegel: *Decision-making: An Experimental Approach.* Stanford: Stanford University Press, 1957.

[8] Debreu, G.: "Topological Methods in Cardinal Utility Theory," *Mathematical Methods in the Social Sciences*, ed. by K. J. Arrow, S. Karlin, and P. Suppes. Stanford: Stanford University Press, 1960, pp. 16–26.

[9] Edwards, W.: "Subjective Probabilities Inferred from Decisions," *Psychological Review*, 69 (1962), 109–135.

[10] Ellsberg, D.: "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics*, 75 (1961), 643–669.

[11] Fellner, W.: "Distortion of Subjective Probabilities as a Reaction to Uncertainty," *Quarterly Journal of Economics*, 75 (1961), 670–690.

[12] ———: *Probability and Profit—A Study of Economic Behavior Along Bayesian Lines.* Homewood, Illinois: Richard D. Irwin, 1965.

[13] Fishburn, P. C.: "Mean-Risk Analysis with Risk Associated with Below-Target Returns," *American Economic Review*, 67 (1977), 116–126.

[14] Fishburn, P. C., and G. A. Kochenberger: "Two-Piece von Neumann-Morgenstern Utility Functions," forthcoming.

[15] Friedman, M., and L. J. Savage: "The Utility Analysis of Choices Involving Risks," *Journal of Political Economy*, 56 (1948), 279–304.

[16] Fuchs, V. R.: "From Bismark to Woodcock: The "Irrational" Pursuit of National Health Insurance," *Journal of Law and Economics*, 19 (1976), 347–359.

[17] Galanter, E., and P. Pliner: "Cross-Modality Matching of Money Against Other Continua," in *Sensation and Measurement*, ed. by H. R. Moskowitz et al. Dordrecht, Holland: Reidel, 1974, pp. 65–76.

[18] GRAYSON, C. J.: *Decisions under Uncertainty: Drilling Decisions by Oil and Gas Operators.* Cambridge, Massachusetts: Graduate School of Business, Harvard University, 1960.

[19] GREEN, P. E.: "Risk Attitudes and Chemical Investment Decisions," *Chemical Engineering Progress*, 59 (1963), 35–40.

[20] GRETHER, D. M., AND C. R. PLOTT: "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, forthcoming.

[21] HALTER, A. N., AND G. W. DEAN: *Decisions under Uncertainty.* Cincinnati: South Western Publishing Co., 1971.

[22] HANSSON, B.: "The Appropriateness of the Expected Utility Model," *Erkenntnis*, 9 (1975), 175–194.

[23] HELSON, H.: *Adaptation-Level Theory.* New York: Harper, 1964.

[24] KEENEY, R. L., AND H. RAIFFA: *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* New York: Wiley, 1976.

[25] KRANTZ, D. H., D. R. LUCE, P. SUPPES, AND A. TVERSKY: *Foundations of Measurement.* New York: Academic Press, 1971.

[26] KUNREUTHER, H., R. GINSBERG, L. MILLER, P. SAGI, P. SLOVIC, B. BORKAN, AND N. KATZ: *Disaster Insurance Protection: Public Policy Lessons.* New York: Wiley, 1978.

[27] LICHTENSTEIN, S, AND P. SLOVIC: "Reversal of Preference Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 89 (1971), 46–55.

[28] MACCRIMMON, K. R., AND S. LARSSON: "Utility Theory: Axioms versus Paradoxes," in *Expected Utility Hypothesis and the Allais Paradox*, ed. by M. Allais and O. Hagen, forthcoming in *Theory and Decision.*

[29] MARKOWITZ, H.: "The Utility of Wealth," *Journal of Political Economy*, 60 (1952), 151–158.

[30] ———: *Portfolio Selection.* New York: Wiley, 1959.

[31] MCGLOTHLIN, W. H.: "Stability of Choices among Uncertain Alternatives," *American Journal of Psychology*, 69 (1956), 604–615.

[32] MOSTELLER, F., AND P. NOGEE: "An Experimental Measurement of Utility," *Journal of Political Economy*, 59 (1951), 371–404.

[33] PRATT, J. W.: "Risk Aversion in the Small and in the Large," *Econometrica*, 32 (1964), 122–136.

[34] RAIFFA H.: *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* Reading, Massachusetts: Addison-Wesley, 1968.

[35] ROSKIES, R.: "A Measurement Axiomatization for an Essentially Multiplicative Representation of Two Factors," *Journal of Mathematical Psychology*, 2 (1965), 266–276.

[36] SAVAGE, L. J.: *The Foundations of Statistics.* New York: Wiley, 1954.

[37] SLOVIC, P., B. FISCHHOFF, S. LICHTENSTEIN, B. CORRIGAN, AND B. COOMBS: "Preference for Insuring Against Probable Small Losses: Insurance Implications," *Journal of Risk and Insurance*, 44 (1977), 237–258.

[38] SLOVIC, P., AND A. TVERSKY: "Who Accepts Savage's Axiom?," *Behavioral Science*, 19 (1974), 368–373.

[39] SPETZLER, C. S.: "The Development of Corporate Risk Policy for Capital Investment Decisions," *IEEE Transactions on Systems Science and Cybernetics*, SSC-4 (1968), 279–300.

[40] SWALM, R. O.: "Utility Theory—Insights into Risk Taking," *Harvard Business Review*, 44 (1966), 123–136.

[41] TOBIN, J.: "Liquidity Preferences as Behavior Towards Risk," *Review of Economic Studies*, 26 (1958), 65–86.

[42] TVERSKY, A.: "Additivity, Utility, and Subjective Probability," *Journal of Mathematical Psychology*, 4 (1967), 175–201.

[43] ——: "Intransitivity of Preferences," *Psychological Review*, 76 (1969), 31–48.

[44] ———: "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79 (1972), 281–299.

[45] TVERSKY, A., AND D. KAHNEMAN: "Judgment under Uncertainty: Heuristics and Biases," *Science*, 185 (1974), 1124–1131.

[46] VAN DAM, C.: "Another Look at Inconsistency in Financial Decision-Making," presented at the Seminar on Recent Research in Finance and Monetary Economics, Cergy-Pontoise, March, 1975.

[47] VON NEUMANN, J., AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, Princeton: Princeton University Press, 1944.

[48] WILLIAMS, A. C.: "Attitudes toward Speculative Risks as an Indicator of Attitudes toward Pure Risks," *Journal of Risk and Insurance*, 33 (1966), 577–586.

Psychological Review

VOLUME 84 NUMBER 4 JULY 1977

Features of Similarity

Amos Tversky
Hebrew University
Jerusalem, Israel

The metric and dimensional assumptions that underlie the geometric representation of similarity are questioned on both theoretical and empirical grounds. A new set-theoretical approach to similarity is developed in which objects are represented as collections of features, and similarity is described as a feature-matching process. Specifically, a set of qualitative assumptions is shown to imply the contrast model, which expresses the similarity between objects as a linear combination of the measures of their common and distinctive features. Several predictions of the contrast model are tested in studies of similarity with both semantic and perceptual stimuli. The model is used to uncover, analyze, and explain a variety of empirical phenomena such as the role of common and distinctive features, the relations between judgments of similarity and difference, the presence of asymmetric similarities, and the effects of context on judgments of similarity. The contrast model generalizes standard representations of similarity data in terms of clusters and trees. It is also used to analyze the relations of prototypicality and family resemblance.

Similarity plays a fundamental role in theories of knowledge and behavior. It serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations. Indeed, the concept of similarity is ubiquitous in psychological theory. It underlies the accounts of stimulus and response generalization in learning, it is employed to explain errors in memory and pattern recognition, and it is central to the analysis of connotative meaning.

Similarity or dissimilarity data appear in different forms: ratings of pairs, sorting of objects, communality between associations, errors of substitution, and correlation between occurrences. Analyses of these data attempt to explain the observed similarity relations and to capture the underlying structure of the objects under study.

The theoretical analysis of similarity relations has been dominated by geometric models. These models represent objects as points in some coordinate space such that the observed dissimilarities between objects correspond to the metric distances between the respective points. Practically all analyses of proximity data have been metric in nature, although some (e.g., hierarchical clustering) yield tree-like structures rather than dimensionally organized spaces. However, most theoretical and empirical analyses of similarity assume that objects can be adequately represented as points in some coordinate space and that dissimilarity behaves like a metric dis-

This paper benefited from fruitful discussions with Y. Cohen, I. Gati, D. Kahneman, L. Sjöberg, and S. Sattath.

Requests for reprints should be sent to Amos Tversky, Department of Psychology, Hebrew University, Jerusalem, Israel.

tance function. Both dimensional and metric assumptions are open to question.

It has been argued by many authors that dimensional representations are appropriate for certain stimuli (e.g., colors, tones) but not for others. It seems more appropriate to represent faces, countries, or personalities in terms of many qualitative features than in terms of a few quantitative dimensions. The assessment of similarity between such stimuli, therefore, may be better described as a comparison of features rather than as the computation of metric distance between points.

A metric distance function, δ, is a scale that assigns to every pair of points a nonnegative number, called their distance, in accord with the following three axioms:

Minimality:

$$\delta(a,b) \geq \delta(a,a) = 0.$$

Symmetry:

$$\delta(a,b) = \delta(b,a).$$

The triangle inequality:

$$\delta(a,b) + \delta(b,c) \geq \delta(a,c).$$

To evaluate the adequacy of the geometric approach, let us examine the validity of the metric axioms when δ is regarded as a measure of dissimilarity. The minimality axiom implies that the similarity between an object and itself is the same for all objects. This assumption, however, does not hold for some similarity measures. For example, the probability of judging two identical stimuli as "same" rather than "different" is not constant for all stimuli. Moreover, in recognition experiments the off-diagonal entries often exceed the diagonal entries; that is, an object is identified as another object more frequently than it is identified as itself. If identification probability is interpreted as a measure of similarity, then these observations violate minimality and are, therefore, incompatible with the distance model.

Similarity has been viewed by both philosophers and psychologists as a prime example of a symmetric relation. Indeed, the assumption of symmetry underlies essentially all theoretical treatments of similarity. Contrary to this tradition, the present paper provides empirical evidence for asymmetric similarities and argues that similarity should not be treated as a symmetric relation.

Similarity judgments can be regarded as extensions of similarity statements, that is, statements of the form "a is like b." Such a statement is directional; it has a subject, a, and a referent, b, and it is not equivalent in general to the converse similarity statement "b is like a." In fact, the choice of subject and referent depends, at least in part, on the relative salience of the objects. We tend to select the more salient stimulus, or the prototype, as a referent, and the less salient stimulus, or the variant, as a subject. We say "the portrait resembles the person" rather than "the person resembles the portrait." We say "the son resembles the father" rather than "the father resembles the son." We say "an ellipse is like a circle," not "a circle is like an ellipse," and we say "North Korea is like Red China" rather than "Red China is like North Korea."

As will be demonstrated later, this asymmetry in the *choice* of similarity statements is associated with asymmetry in *judgments* of similarity. Thus, the judged similarity of North Korea to Red China exceeds the judged similarity of Red China to North Korea. Likewise, an ellipse is more similar to a circle than a circle is to an ellipse. Apparently, the direction of asymmetry is determined by the relative salience of the stimuli; the variant is more similar to the prototype than vice versa.

The directionality and asymmetry of similarity relations are particularly noticeable in similies and metaphors. We say "Turks fight like tigers" and not "tigers fight like Turks." Since the tiger is renowned for its fighting spirit, it is used as the referent rather than the subject of the simile. The poet writes "my love is as deep as the ocean," not "the ocean is as deep as my love," because the ocean epitomizes depth. Sometimes both directions are used but they carry different meanings. "A man is like a tree" implies that man has roots; "a tree is like a man" implies that the tree has a life history. "Life is like a play" says that people play roles. "A play is like life" says that a play can capture the essential elements of human life. The relations between the interpretation of metaphors and the assessment of similarity are briefly discussed in the final section.

The triangle inequality differs from minimality and symmetry in that it cannot be formulated in ordinal terms. It asserts that one distance must be smaller than the sum of two others, and hence it cannot be readily refuted with ordinal or even interval data. However, the triangle inequality implies that if a is quite similar to b, and b is quite similar to c, then a and c cannot be very dissimilar from each other. Thus, it sets a lower limit to the similarity between a and c in terms of the similarities between a and b and between b and c. The following example (based on William James) casts some doubts on the psychological validity of this assumption. Consider the similarity between countries: Jamaica is similar to Cuba (because of geographical proximity); Cuba is similar to Russia (because of their political affinity); but Jamaica and Russia are not similar at all.

This example shows that similarity, as one might expect, is not transitive. In addition, it suggests that the perceived distance of Jamaica to Russia exceeds the perceived distance of Jamaica to Cuba, plus that of Cuba to Russia —contrary to the triangle inequality. Although such examples do not necessarily refute the triangle inequality, they indicate that it should not be accepted as a cornerstone of similarity models.

It should be noted that the metric axioms, by themselves, are very weak. They are satisfied, for example, by letting $\delta(a,b) = 0$ if $a = b$, and $\delta(a,b) = 1$ if $a \neq b$. To specify the distance function, additional assumptions are made (e.g., intradimensional subtractivity and interdimensional additivity) relating the dimensional structure of the objects to their metric distances. For an axiomatic analysis and a critical discussion of these assumptions, see Beals, Krantz, and Tversky (1968), Krantz and Tversky (1975), and Tversky and Krantz (1970).

In conclusion, it appears that despite many fruitful applications (see e.g., Carroll & Wish, 1974; Shepard, 1974), the geometric approach to the analysis of similarity faces several difficulties. The applicability of the dimensional assumption is limited, and the metric axioms are questionable. Specifically, minimality is somewhat problematic, symmetry is apparently false, and the triangle inequality is hardly compelling.

The next section develops an alternative theoretical approach to similarity, based on feature matching, which is neither dimensional nor metric in nature. In subsequent sections this approach is used to uncover, analyze, and explain several empirical phenomena, such as the role of common and distinctive features, the relations between judgments of similarity and difference, the presence of asymmetric similarities, and the effects of context on similarity. Extensions and implications of the present development are discussed in the final section.

Feature Matching

Let $\Delta = \{a,b,c,\ldots\}$ be the domain of objects (or stimuli) under study. Assume that each object in Δ is represented by a set of features or attributes, and let A,B,C denote the sets of features associated with the objects a,b,c, respectively. The features may correspond to components such as eyes or mouth; they may represent concrete properties such as size or color; and they may reflect abstract attributes such as quality or complexity. The characterization of stimuli as feature sets has been employed in the analysis of many cognitive processes such as speech perception (Jakobson, Fant, & Halle, 1961), pattern recognition (Neisser, 1967), perceptual learning (Gibson, 1969), preferential choice (Tversky, 1972), and semantic judgment (Smith, Shoben, & Rips, 1974).

Two preliminary comments regarding feature representations are in order. First, it is important to note that our total data base concerning a particular object (e.g., a person, a country, or a piece of furniture) is generally rich in content and complex in form. It includes appearance, function, relation to other objects, and any other property of the object that can be deduced from our general knowledge of the world. When faced with a particular task (e.g., identification or similarity assessment) we extract and compile from our data base a limited list of relevant features on the basis of which we perform the required task. Thus, the representation of an object as a col-

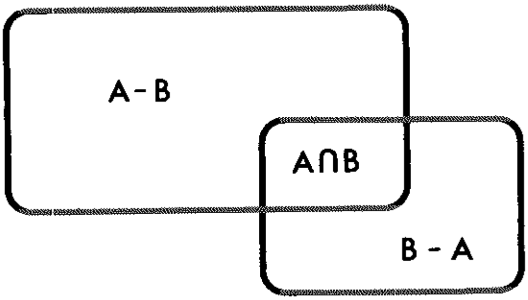


Figure 1. A graphical illustration of the relation between two feature sets.

lection of features is viewed as a product of a prior process of extraction and compilation.

Second, the term *feature* usually denotes the value of a binary variable (e.g., voiced vs. voiceless consonants) or the value of a nominal variable (e.g., eye color). Feature representations, however, are not restricted to binary or nominal variables; they are also applicable to ordinal or cardinal variables (i.e., dimensions). A series of tones that differ only in loudness, for example, could be represented as a sequence of nested sets where the feature set associated with each tone is included in the feature sets associated with louder tones. Such a representation is isomorphic to a directional unidimensional structure. A nondirectional unidimensional structure (e.g., a series of tones that differ only in pitch) could be represented by a chain of overlapping sets. The set-theoretical representation of qualitative and quantitative dimensions has been investigated by Restle (1959).

Let $s(a,b)$ be a measure of the similarity of a to b defined for all distinct a, b in Δ. The scale s is treated as an ordinal measure of similarity. That is, $s(a,b) > s(c,d)$ means that a is more similar to b than c is to d. The present theory is based on the following assumptions.

1. *Matching:*

$$s(a,b) = F(A \cap B, A - B, B - A).$$

The similarity of a to b is expressed as a function F of three arguments: $A \cap B$, the features that are common to both a and b; $A - B$, the features that belong to a but not to b; $B - A$, the features that belong to b but not to a. A schematic illustration of these components is presented in Figure 1.

2. *Monotonicity:*

$$s(a,b) \geq s(a,c)$$

whenever

$$A \cap B \supset A \cap C, \quad A - B \subset A - C,$$

and

$$B - A \subset C - A.$$

Moreover, the inequality is strict whenever either inclusion is proper.

That is, similarity increases with addition of common features and/or deletion of distinctive features (i.e., features that belong to one object but not to the other). The monotonicity axiom can be readily illustrated with block letters if we identify their features with the component (straight) lines. Under this assumption, E should be more similar to F than to I because E and F have more common features than E and I. Furthermore, I should be more similar to F than to E because I and F have fewer distinctive features than I and E.

Any function F satisfying Assumptions 1 and 2 is called a *matching function.* It measures the degree to which two objects—viewed as sets of features—match each other. In the present theory, the assessment of similarity is described as a feature-matching process. It is formulated, therefore, in terms of the set-theoretical notion of a matching function rather than in terms of the geometric concept of distance.

In order to determine the functional form of the matching function, additional assumptions about the similarity ordering are introduced. The major assumption of the theory (independence) is presented next; the remaining assumptions and the proof of the representation theorem are presented in the Appendix. Readers who are less interested in formal theory can skim or skip the following paragraphs up to the discussion of the representation theorem.

Let Φ denote the set of all features associated with the objects of Δ, and let X,Y,Z,...etc. denote collections of features (i.e., subsets of Φ). The expression $F(X,Y,Z)$ is defined whenever there exists a, b in Δ such that $A \cap B = X$,

$A - B = Y$, and $B - A = Z$, whence $s(a,b) = F(A \cap B, A - B, B - A) = F(X,Y,Z)$. Next, define $V \simeq W$ if one or more of the following hold for some X,Y,Z: $F(V,Y,Z) = F(W,Y,Z)$, $F(X,V,Z) = F(X,W,Z)$, $F(X,Y,V) = F(X,Y,W)$.

The pairs (a,b) and (c,d) are said to *agree* on one, two, or three components, respectively, whenever one, two, or three of the following hold: $(A \cap B) \simeq (C \cap D)$, $(A - B) \simeq (C - D)$, $(B - A) \simeq (D - C)$.

3. *Independence:* Suppose the pairs (a,b) and (c,d), as well as the pairs (a′,b′) and (c′,d′), agree on the same two components, while the pairs (a,b) and (a′,b′), as well as the pairs (c,d) and (c′,d′), agree on the remaining (third) component. Then

$$s(a,b) \geq s(a',b') \quad \text{iff} \quad s(c,d) \geq s(c',d').$$

To illustrate the force of the independence axiom consider the stimuli presented in Figure 2, where

$$A \cap B = C \cap D = \text{round profile} = X,$$
$$A' \cap B' = C' \cap D' = \text{sharp profile} = X',$$
$$A - B = C - D = \text{smiling mouth} = Y,$$
$$A' - B' = C' - D' = \text{frowning mouth} = Y',$$
$$B - A = B' - A' = \text{straight eyebrow} = Z,$$
$$D - C = D' - C' = \text{curved eyebrow} = Z'.$$

By independence, therefore,

$$\begin{aligned} s(a,b) &= F(A \cap B, A - B, B - A) \\ &= F(X,Y,Z) \geq F(X',Y',Z) \\ &= F(A' \cap B', A' - B', B' - A') \\ &= s(a',b') \end{aligned}$$

if and only if

$$\begin{aligned} s(c,d) &= F(C \cap D, C - D, D - C) \\ &= F(X,Y,Z') \geq F(X',Y',Z') \\ &= F(C' \cap D', C' - D', D' - C') \\ &= s(c',d'). \end{aligned}$$

Thus, the ordering of the joint effect of any two components (e.g., X,Y vs. X′,Y′) is independent of the fixed level of the third factor (e.g., Z or Z′).

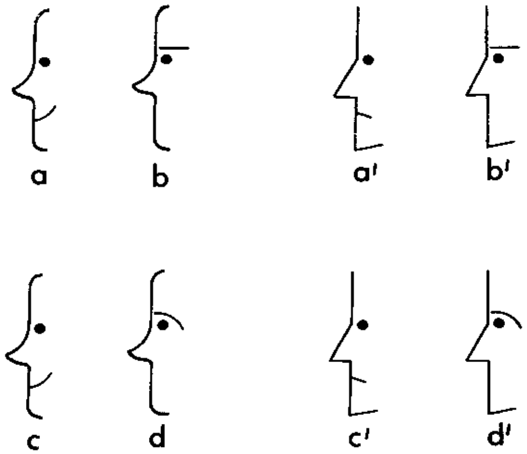


Figure 2. An illustration of independence.

It should be emphasized that any test of the axioms presupposes an interpretation of the features. The independence axiom, for example, may hold in one interpretation and fail in another. Experimental tests of the axioms, therefore, test jointly the adequacy of the interpretation of the features and the empirical validity of the assumptions. Furthermore, the above examples should not be taken to mean that stimuli (e.g., block letters, schematic faces) can be properly characterized in terms of their components. To achieve an adequate feature representation of visual forms, more global properties (e.g., symmetry, connectedness) should also be introduced. For an interesting discussion of this problem, in the best tradition of Gestalt psychology, see Goldmeier (1972; originally published in 1936).

In addition to matching (1), monotonicity (2), and independence (3), we also assume solvability (4), and invariance (5). Solvability requires that the feature space under study be sufficiently rich that certain (similarity) equations can be solved. Invariance ensures that the equivalence of intervals is preserved across factors. A rigorous formulation of these assumptions is given in the Appendix, along with a proof of the following result.

Representation theorem. Suppose Assumptions 1, 2, 3, 4, and 5 hold. Then there exist a similarity scale S and a nonnegative scale f such that for all a,b,c,d in Δ,

(i). $S(a,b) \geq S(c,d)$ iff $s(a,b) \geq s(c,d)$;

(ii). $S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$, for some $\theta,\alpha,\beta \geq 0$;

(iii). f and S are interval scales.

The theorem shows that under Assumptions 1–5, there exists an interval similarity scale S that preserves the observed similarity order and expresses similarity as a linear combination, or a contrast, of the measures of the common and the distinctive features. Hence, the representation is called the *contrast model*. In parts of the following development we also assume that f satisfies feature additivity. That is, $f(X \cup Y) = f(X) + f(Y)$ whenever X and Y are disjoint, and all three terms are defined[1].

Note that the contrast model does not define a single similarity scale, but rather a family of scales characterized by different values of the parameters θ, α, and β. For example, if $\theta = 1$ and α and β vanish, then $S(a,b) = f(A \cap B)$; that is, the similarity between objects is the measure of their common features. If, on the other hand, $\alpha = \beta = 1$ and θ vanishes then $-S(a,b) = f(A - B) + f(B - A)$; that is, the dissimilarity between objects is the measure of the symmetric difference between the respective feature sets. Restle (1961) has proposed these forms as models of similarity and psychological distance, respectively. Note that in the former model ($\theta = 1$, $\alpha = \beta = 0$), similarity between objects is determined only by their common features, whereas in the latter model ($\theta = 0, \alpha = \beta = 1$), it is determined by their distinctive features only. The contrast model expresses similarity between objects as a weighted difference of the measures of their common and distinctive features, thereby allowing for a variety of similarity relations over the same domain.

The major constructs of the present theory are the contrast rule for the assessment of similarity, and the scale f, which reflects the salience or prominence of the various features. Thus, f measures the contribution of any particular (common or distinctive) feature to the similarity between objects. The scale value f(A) associated with stimulus a is regarded, therefore, as a measure of the overall salience of that stimulus. The factors that contribute to the salience of a stimulus include intensity, frequency, familiarity, good form, and informational content. The manner in which the scale f and the parameters (θ,α,β) depend on the context and the task are discussed in the following sections.

Let us recapitulate what is assumed and what is proven in the representation theorem. We begin with a set of objects, described as collections of features, and a similarity ordering which is assumed to satisfy the axioms of the present theory. From these assumptions, we derive a measure f on the feature space and prove that the similarity ordering of object pairs coincides with the ordering of their contrasts, defined as linear combinations of the respective common and distinctive features. Thus, the measure f and the contrast model are derived from qualitative axioms regarding the similarity of objects.

The nature of this result may be illuminated by an analogy to the classical theory of decision under risk (von Neumann & Morgenstern, 1947). In that theory, one starts with a set of prospects, characterized as probability distributions over some consequence space, and a preference order that is assumed to satisfy the axioms of the theory. From these assumptions one derives a utility scale on the consequence space and proves that the preference order between prospects coincides with the order of their expected utilities. Thus, the utility scale and the expectation principle are derived from qualitative assumptions about preferences. The present theory of similarity differs from the expected-utility model in that the characterization of objects as feature sets is perhaps more problematic than the characterization of uncertain options as probability distributions. Furthermore, the axioms of utility theory are proposed as (normative) principles of rational behavior, whereas the axioms of the present theory are intended to be descriptive rather than prescriptive.

The contrast model is perhaps the simplest form of a matching function, yet it is not the only form worthy of investigation. Another

[1] To derive feature additivity from qualitative assumptions, we must assume the axioms of an extensive structure and the compatibility of the extensive and the conjoint scales; see Krantz et al. (1971, Section 10.7).

matching function of interest is the *ratio model*,

$$S(a,b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)}, \quad \alpha, \beta \geq 0,$$

where similarity is normalized so that S lies between 0 and 1. The ratio model generalizes several set-theoretical models of similarity proposed in the literature. If $\alpha = \beta = 1$, S(a,b) reduces to $f(A \cap B)/f(A \cup B)$ (see Gregson, 1975, and Sjöberg, 1972). If $\alpha = \beta = \frac{1}{2}$, S(a,b) equals $2f(A \cap B)/(f(A) + f(B))$ (see Eisler & Ekman, 1959). If $\alpha = 1$ and $\beta = 0$, S(a,b) reduces to $f(A \cap B)/f(A)$ (see Bush & Mosteller, 1951). The present framework, therefore, encompasses a wide variety of similarity models that differ in the form of the matching function F and in the weights assigned to its arguments.

In order to apply and test the present theory in any particular domain, some assumptions about the respective feature structure must be made. If the features associated with each object are explicitly specified, we can test the axioms of the theory directly and scale the features according to the contrast model. This approach, however, is generally limited to stimuli (e.g., schematic faces, letters, strings of symbols) that are constructed from a fixed feature set. If the features associated with the objects under study cannot be readily specified, as is often the case with natural stimuli, we can still test several predictions of the contrast model which involve only general qualitative assumptions about the feature structure of the objects. Both approaches were employed in a series of experiments conducted by Itamar Gati and the present author. The following three sections review and discuss our main findings, focusing primarily on the test of qualitative predictions. A more detailed description of the stimuli and the data are presented in Tversky and Gati (in press).

Asymmetry and Focus

According to the present analysis, similarity is not necessarily a symmetric relation. Indeed, it follows readily (from either the contrast or the ratio model) that

$$\begin{aligned} s(a,b) = s(b,a) \quad \text{iff} \quad & \alpha f(A - B) + \beta f(B - A) \\ & = \alpha f(B - A) + \beta f(A - B) \\ \text{iff} \quad & (\alpha - \beta) f(A - B) = (\alpha - \beta) f(B - A). \end{aligned}$$

Hence, $s(a,b) = s(b,a)$ if either $\alpha = \beta$, or $f(A - B) = f(B - A)$, which implies $f(A) = f(B)$, provided feature additivity holds. Thus, symmetry holds whenever the objects are equal in measure $(f(A) = f(B))$ or the task is nondirectional $(\alpha = \beta)$. To interpret the latter condition, compare the following two forms:

(i). Assess the degree to which a and b are similar to each other.
(ii). Assess the degree to which a is similar to b.

In (i), the task is formulated in a nondirectional fashion; hence it is expected that $\alpha = \beta$ and $s(a,b) = s(b,a)$. In (ii), on the other hand, the task is directional, and hence α and β may differ and symmetry need not hold.

If s(a,b) is interpreted as the degree to which a is similar to b, then a is the subject of the comparison and b is the referent. In such a task, one naturally focuses on the subject of the comparison. Hence, the features of the subject are weighted more heavily than the features of the referent (i.e., $\alpha > \beta$). Consequently, similarity is reduced more by the distinctive features of the subject than by the distinctive features of the referent. It follows readily that whenever $\alpha > \beta$,

$$s(a,b) > s(b,a) \quad \text{iff} \quad f(B) > f(A).$$

Thus, the focusing hypothesis (i.e., $\alpha > \beta$) implies that the direction of asymmetry is determined by the relative salience of the stimuli so that the less salient stimulus is more similar to the salient stimulus than vice versa. In particular, the variant is more similar to the prototype than the prototype is to the variant, because the prototype is generally more salient than the variant.

Similarity of Countries

Twenty-one pairs of countries served as stimuli. The pairs were constructed so that one element was more prominent than the other (e.g., Red China–North Vietnam, USA–Mexico, Belgium–Luxemburg). To verify this relation, we asked a group of 69 subjects[2] to select in

[2] The subjects in all our experiments were Israeli college students, ages 18–28. The material was presented in booklets and administered in a group setting.

each pair the country they regarded as more prominent. The proportion of subjects that agreed with the a priori ordering exceeded $\frac{2}{3}$ for all pairs except one. A second group of 69 subjects was asked to choose which of two phrases they preferred to use: "country a is similar to country b," or "country b is similar to country a." In all 21 cases, most of the subjects chose the phrase in which the less prominent country served as the subject and the more prominent country as the referent. For example, 66 subjects selected the phrase "North Korea is similar to Red China" and only 3 selected the phrase "Red China is similar to North Korea." These results demonstrate the presence of marked asymmetries in the choice of similarity statements, whose direction coincides with the relative prominence of the stimuli.

To test for asymmetry in direct judgments of similarity, we presented two groups of 77 subjects each with the same list of 21 pairs of countries and asked subjects to rate their similarity on a 20-point scale. The only difference between the two groups was the order of the countries within each pair. For example, one group was asked to assess "the degree to which the USSR is similar to Poland," whereas the second group was asked to assess "the degree to which Poland is similar to the USSR." The lists were constructed so that the more prominent country appeared about an equal number of times in the first and second positions.

For any pair (p,q) of stimuli, let p denote the more prominent element, and let q denote the less prominent element. The average s(q,p) was significantly higher than the average s(p,q) across all subjects and pairs: t test for correlated samples yielded $t(20) = 2.92$, $p < .01$. To obtain a statistical test based on individual data, we computed for each subject a directional asymmetry score defined as the average similarity for comparisons with a prominent referent, that is, s(q,p), minus the average similarity for comparisons with a prominent subject, s(p,q). The average difference was significantly positive: $t(153) = 2.99$, $p < .01$.

The above study was repeated using judgments of difference instead of judgments of similarity. Two groups of 23 subjects each participated in this study. They received the same list of 21 pairs except that one group was asked to judge the degree to which country a differed from country b, denoted d(a,b), whereas the second group was asked to judge the degree to which country b was different from country a, denoted d(b,a). If judgments of difference follow the contrast model, and $\alpha > \beta$, then we expect the prominent stimulus p to differ from the less prominent stimulus q more than q differs from p; that is, d(p,q) > d(q,p). This hypothesis was tested using the same set of 21 pairs of countries and the prominence ordering established earlier. The average d(p,q), across all subjects and pairs, was significantly higher than the average d(q,p): t test for correlated samples yielded $t(20) = 2.72$, $p < .01$. Furthermore, the average asymmetry score, computed as above for each subject, was significantly positive, $t(45) = 2.24$, $p < .05$.

Similarity of Figures

A major determinant of the salience of geometric figures is goodness of form. Thus, a "good figure" is likely to be more salient than a "bad figure," although the latter is generally more complex. However, when two figures are roughly equivalent with respect to goodness of form, the more complex figure is likely to be more salient. To investigate these hypotheses and to test the asymmetry prediction, two sets of eight pairs of geometric figures were constructed. In the first set, one figure in each pair (denoted p) had better form than the other (denoted q). In the second set, the two figures in each pair were roughly matched in goodness of form, but one figure (denoted p) was richer or more complex than the other (denoted q). Examples of pairs of figures from each set are presented in Figure 3.

A group of 69 subjects was presented with the entire list of 16 pairs of figures, where the two elements of each pair were displayed side by side. For each pair, the subjects were asked to indicate which of the following two statements they preferred to use: "The left figure is similar to the right figure," or "The right figure is similar to the left figure." The positions of the stimuli were randomized so that p and q appeared an equal number of times on the left and on the right. The results showed that in each one of the pairs, most of the subjects selected the form "q is similar to p." Thus, the more salient stimulus was generally chosen as the referent rather than the subject of similarity statements.

To test for asymmetry in judgments of similarity, we presented two groups of 67 subjects each with the same 16 pairs of figures and asked the subjects to rate (on a 20-point scale) the degree to which the figure on the left was similar to the figure on the right. The two groups received identical booklets, except that the left and right positions of the figures in each pair were reversed. The results showed that the average s(q,p) across all subjects and pairs was significantly higher than the average s(p,q). A t test for correlated samples yielded $t(15) = 2.94$, $p < .01$. Furthermore, in both sets the average asymmetry scores, computed as above for each subject, were significantly positive: In the first set $t(131) = 2.96$, $p < .01$, and in the second set $t(131) = 2.79$, $p < .01$.

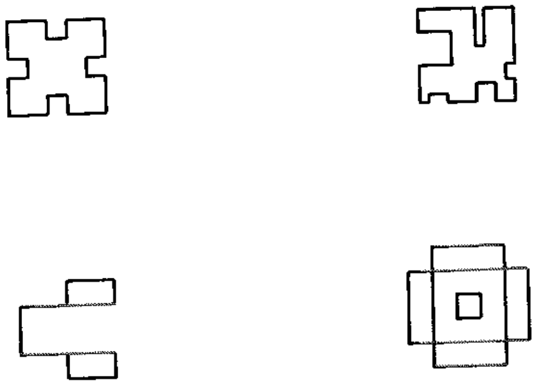


Figure 3. Examples of pairs of figures used to test the prediction of asymmetry. The top two figures are examples of a pair (from the first set) that differs in goodness of form. The bottom two are examples of a pair (from the second set) that differs in complexity.

Similarity of Letters

A common measure of similarity between stimuli is the probability of confusing them in a recognition or an identification task: The more similar the stimuli, the more likely they are to be confused. While confusion probabilities are often asymmetric (i.e., the probability of confusing a with b is different from the probability of confusing b with a), this effect is typically attributed to a response bias. To eliminate this interpretation of asymmetry, one could employ an experimental task where the subject merely indicates whether the two stimuli presented to him (sequentially or simultaneously) are identical or not. This procedure was employed by Yoav Cohen and the present author in a study of confusion among block letters.

The following eight block letters served as stimuli: ⌈, ⊏, ⊓, □, ⊧, ⋿, ⍓, ⊟. All pairs of letters were displayed on a cathode-ray tube, side by side, on a noisy background. The letters were presented sequentially, each for approximately 1 msec. The right letter always followed the left letter with an interval of 630 msec in between. After each presentation the subject pressed one of two keys to indicate whether the two letters were identical or not.

A total of 32 subjects participated in the experiment. Each subject was tested individually. On each trial, one letter (known in advance) served as the standard. For one half of the subjects the standard stimulus always appeared on the left, and for the other half of the subjects the standard always appeared on the right. Each one of the eight letters served as a standard. The trials were blocked into groups of 10 pairs in which the standard was paired once with each of the other letters and three times with itself. Since each letter served as a standard in one block, the entire design consisted of eight blocks of 10 trials each. Every subject was presented with three replications of the entire design (i.e., 240 trials). The order of the blocks in each design and the order of the letters within each block were randomized.

According to the present analysis, people compare the variable stimulus, which serves the role of the subject, to the standard (i.e., the referent). The choice of standard, therefore, determines the directionality of the comparison. A natural partial ordering of the letters with respect to prominence is induced by the relation of inclusion among letters. Thus, one letter is assumed to have a larger measure than another if the former includes the latter. For example, ⋿ includes ⊧ and ⌈ but not □. For all 19 pairs in which one letter includes the other, let p denote the more prominent letter and q denote the less prominent letter. Furthermore, let s(a,b) denote the percentage of times that the subject judged the variable stimulus a to be the same as the standard b.

It follows from the contrast model, with $\alpha > \beta$, that the proportion of "same" responses should be larger when the variable is included in the standard than when the standard is included in the variable, that is, s(q,p) > s(p,q). This prediction was borne out by the data. The average s(q,p) across all subjects and trials was 17.1%, whereas the average s(p,q) across all subjects and trials was 12.4%. To obtain a statistical test, we computed for each subject the difference between s(q,p) and s(p,q) across all trials. The difference was significantly positive, $t(31) = 4.41$, $p < .001$. These results demonstrate that the prediction of directional asymmetry derived from the contrast model applies to confusion data and not merely to rated similarity.

Similarity of Signals

Rothkopf (1957) presented 598 subjects with all ordered pairs of the 36 Morse Code signals and asked them to indicate whether the two signals in each pair were the same or not. The pairs were presented in a randomized order without a fixed standard. Each subject judged about one fourth of all pairs.

Let s(a,b) denote the percentage of "same" responses to the ordered pair (a,b), i.e., the percentage of subjects that judged the first signal a to be the same as the second signal b. Note that a and b refer here to the first and second signal, and not to the variable and the standard as in the previous section. Obviously, Morse Code signals are partially ordered according to temporal length. For any pair of signals that differ in temporal length, let p and q denote, respectively, the longer and shorter element of the pair.

From the total of 555 comparisons between signals of different length, reported in Rothkopf (1957), s(q,p) exceeds s(p,q) in 336 cases, s(p,q) exceeds s(q,p) in 181 cases, and s(q,p) equals s(p,q) in 38 cases, $p < .001$, by sign test. The average difference between s(q,p) and s(p,q) across all pairs is 3.3%, which is also highly significant. A t test for correlated samples yields $t(554) = 9.17$, $p < .001$.

The asymmetry effect is enhanced when we consider only those comparisons in which one signal is a proper subsequence of the other. (For example, ·· is a subsequence of ·-- as well as of --·). From a total of 195 comparisons of this type, s(q,p) exceeds s(p,q) in 128 cases, s(p,q) exceeds s(q,p) in 55 cases, and s(q,p) equals s(p,q) in 12 cases, $p < .001$ by sign test. The average difference between s(q,p) and s(p,q) in this case is 4.7%, $t(194) = 7.58$, $p < .001$.

A later study following the same experimental paradigm with somewhat different signals was conducted by Wish (1967). His signals consisted of three tones separated by two silent intervals, where each component (i.e., a tone or a silence) was either short or long. Subjects were presented with all pairs of 32 signals generated in this fashion and judged whether the two members of each pair were the same or not.

The above analysis is readily applicable to Wish's (1967) data. From a total of 386 comparisons between signals of different length, s(q,p) exceeds s(p,q) in 241 cases, s(p,q) exceeds s(q,p) in 117 cases, and s(q,p) equals s(p,q) in 28 cases. These data are clearly asymmetric, $p < .001$ by sign test. The average difference between s(q,p) and s(p,q) is 5.9%, which is also highly significant, $t(385) = 9.23$, $p < .001$.

In the studies of Rothkopf and Wish there is no a priori way to determine the directionality of the comparison, or equivalently to identify the subject and the referent. However, if we accept the focusing hypothesis ($\alpha > \beta$) and the assumption that longer signals are more prominent than shorter ones, then the direction of the observed asymmetry indicates that the first signal serves as the subject that is compared with the second signal that serves the role of the referent. Hence, the directionality of the comparison is determined, according to the present analysis, from the prominence ordering of the stimuli and the observed direction of asymmetry.

Rosch's Data

Rosch (1973, 1975) has articulated and supported the view that perceptual and semantic categories are naturally formed and defined in terms of focal points, or prototypes. Because of the special role of prototypes in the formation of categories, she hypothesized that (i) in sentence frames involving hedges such as "a is essentially b," focal stimuli (i.e., prototypes) appear in the second position; and (ii) the perceived distance from the prototype to the variant is greater than the perceived distance from the variant to the prototype. To test these hypotheses, Rosch (1975) used three stimulus domains: color, line orientation, and number. Prototypical colors were focal (e.g., pure red), while the variants were either nonfocal (e.g., off-red) or less saturated. Vertical, horizontal, and diagonal lines served as prototypes for line orientation, and lines of other angles served as variants. Multiples of 10 (e.g., 10, 50, 100) were taken as prototypical numbers, and other numbers (e.g., 11, 52, 103) were treated as variants.

Hypothesis (i) was strongly confirmed in all three domains. When presented with sentence frames such as "____ is virtually ____," subjects generally placed the prototype in the second blank and the variant in the first. For instance, subjects preferred the sentence "103 is virtually 100" to the sentence "100 is virtually 103." To test hypothesis (ii), one stimulus (the standard) was placed at the origin of a semicircular board, and the subject was instructed to place the second (variable) stimulus on the board so as "to represent his feeling of the distance between that stimulus and the one fixed at the origin." As hypothesized, the measured distance between stimuli was significantly smaller when the prototype, rather than the variant, was fixed at the origin, in each of the three domains.

If focal stimuli are more salient than nonfocal stimuli, then Rosch's findings support the present analysis. The hedging sentences (e.g., "a is roughly b") can be regarded as a particular type of similarity statements. Indeed, the hedges data are in perfect agreement with the choice of similarity statements. Furthermore, the observed asymmetry in distance placement follows from the present analysis of asymmetry and the natural assumptions that the standard and the variable serve, respectively, as referent and subject in the distance-placement task. Thus, the placement of b at distance t from a is interpreted as saying that the (perceived) distance from b to a equals t.

Rosch (1975) attributed the observed asymmetry to the special role of distinct prototypes (e.g., a perfect square or a pure red) in the processing of information. In the present theory, on the other hand, asymmetry is explained by the relative salience of the stimuli. Consequently, it implies asymmetry for pairs that do not include the prototype (e.g., two levels of distortion of the same form). If the concept of prototypicality, however, is interpreted in a relative sense (i.e., a is more prototypical than b) rather than in an absolute sense, then the two interpretations of asymmetry practically coincide.

Discussion

The conjunction of the contrast model and the focusing hypothesis implies the presence of asymmetric similarities. This prediction was confirmed in several experiments of perceptual and conceptual similarity using both judgmental methods (e.g., rating) and behavioral methods (e.g., choice).

The asymmetries discussed in the previous section were observed in *comparative* tasks in which the subject compares two given stimuli to determine their similarity. Asymmetries were also observed in *production* tasks in which the subject is given a single stimulus and asked to produce the most similar response. Studies of pattern recognition, stimulus identification, and word association are all examples of production tasks. A common pattern observed in such studies is that the more salient object occurs more often as a response to the less salient object than vice versa. For example, "tiger" is a more likely associate to "leopard" than "leopard" is to "tiger." Similarly, Garner (1974) instructed subjects to select from a given set of dot patterns one that is similar—but not identical—to a given pattern. His results show that "good" patterns are usually chosen as responses to "bad" patterns and not conversely.

This asymmetry in production tasks has commonly been attributed to the differential availability of responses. Thus, "tiger" is a more likely associate to "leopard" than vice versa, because "tiger" is more common and hence a more available response than "leopard." This account is probably more applicable to situations where the subject must actually produce the response (as in word association or pattern recognition) than to situations where the subject merely selects a response from some specified set (as in Garner's task).

Without questioning the importance of response availability, the present theory suggests another reason for the asymmetry observed in production tasks. Consider the following translation of a production task to a question-and-answer scheme. Question: What is a like? Answer: a is like b. If this interpretation is valid and the given object a serves as a subject rather than as a referent, then the observed asymmetry of production follows from the present theoretical analysis, since $s(a,b) > s(b,a)$ whenever $f(B) > f(A)$.

In summary, it appears that proximity data from both comparative and production tasks reveal significant and systematic asymmetries whose direction is determined by the relative salience of the stimuli. Nevertheless, the symmetry assumption should not be rejected altogether. It seems to hold in many contexts, and it serves as a useful approximation in many others. It cannot be accepted, however, as a universal principle of psychological similarity.

Common and Distinctive Features

In the present theory, the similarity of objects is expressed as a linear combination, or a contrast, of the measures of their common and distinctive features. This section investigates the relative impact of these components and their effect on the relation between the assessments of similarity and difference. The discussion concerns only symmetric tasks, where $\alpha = \beta$, and hence $s(a,b) = s(b,a)$.

Elicitation of Features

The first study employs the contrast model to predict the similarity between objects from features that were produced by the subjects. The following 12 vehicles served as stimuli: bus, car, truck, motorcycle, train, airplane, bicycle, boat, elevator, cart, raft, sled. One group of 48 subjects rated the similarity between all 66 pairs of vehicles on a scale from 1 (no similarity) to 20 (maximal similarity). Following Rosch and Mervis (1975), we instructed a second group of 40 subjects to list the characteristic features of each one of the vehicles. Subjects were given 70 sec to list the features that characterized each vehicle. Different orders of presentation were used for different subjects.

The number of features per vehicle ranged from 71 for airplane to 21 for sled. Altogether, 324 features were listed by the subjects, of which 224 were unique and 100 were shared by two or more vehicles. For every pair of vehicles we counted the number of features that were attributed to both (by at least one subject), and the number of features that were attributed to one vehicle but not to the other. The frequency of subjects that listed each common or distinctive feature was computed.

In order to predict the similarity between vehicles from the listed features, the measures of their common and distinctive features must be defined. The simplest measure is obtained by counting the number of common and distinctive features produced by the subjects. The product-moment correlation between the (average) similarity of objects and the number of their common features was .68. The correlation between the similarity of objects and the number of their distinctive features was −.36. The multiple correlation between similarity and the numbers of common and distinctive features (i.e., the correlation between similarity and the contrast model) was .72.

The counting measure assigns equal weight to all features regardless of their frequency of mention. To take this factor into account, let X_a denote the proportion of subjects who attributed feature X to object a, and let N_X denote the number of objects that share feature X. For any a,b, define the measure of their common features by $f(A \cap B) = \Sigma X_a X_b / N_X$, where the summation is over all X in $A \cap B$, and the measure of their distinctive features by

$$f(A - B) + f(B - A) = \Sigma Y_a + \Sigma Z_b$$

where the summations range over all $Y \epsilon A - B$ and $Z \epsilon B - A$, that is, the distinctive features of a and b, respectively. The correlation between similarity and the above measure of the common features was .84; the correlation between similarity and the above measure of the distinctive features was −.64. The multiple correlation between similarity and the measures of the common and the distinctive features was .87.

Note that the above methods for defining the measure f were based solely on the elicited features and did not utilize the similarity data at all. Under these conditions, a perfect correlation between the two should not be expected because the weights associated with the features are not optimal for the prediction of similarity. A given feature may be frequently mentioned because it is easily labeled or recalled, although it does not have a great impact on similarity, and vice versa. Indeed, when the features were scaled using the additive tree procedure (Sattath & Tversky, in press) in which the measure of the features is derived from the similarities between the objects, the correlation between the data and the model reached .94.

The results of this study indicate that (i) it is possible to elicit from subjects detailed features of semantic stimuli such as vehicles (see Rosch & Mervis, 1975); (ii) the listed features can be used to predict similarity according to the contrast model with a reasonable degree of success; and (iii) the prediction of similarity is improved when frequency of mention and not merely the number of features is taken into account.

Similarity versus Difference

It has been generally assumed that judgments of similarity and difference are complementary; that is, judged difference is a linear function of judged similarity with a slope of −1. This hypothesis has been confirmed in several studies. For example, Hosman and Kuennapas (1972) obtained independent judgments of similarity and difference for all pairs of lowercase letters on a scale from 0 to 100. The product-moment correlation between the judgments was −.98, and the slope of the regression line was −.91. We also collected judgments of similarity and difference for 21 pairs of countries using a 20-point rating scale. The sum of the two judgments for each pair was quite close to 20 in all cases. The product-moment correlation between the ratings was again −.98. This inverse relation between similarity and difference, however, does not always hold.

Naturally, an increase in the measure of the common features increases similarity and decreases difference, whereas an increase in the measure of the distinctive features decreases similarity and increases difference. However, the relative weight assigned to the common and the distinctive features may differ in the two tasks. In the assessment of similarity between objects the subject may attend more to their common features, whereas in the assessment of difference between objects the subject may attend more to their distinctive features. Thus, the relative weight of the common features will be greater in the former task than in the latter task.

Let $d(a,b)$ denote the perceived difference between a and b. Suppose d satisfies the axioms of the present theory with the reverse inequality in the monotonicity axiom, that is, $d(a,b) \leq d(a,c)$ whenever $A \cap B \supset A \cap C$, $A - B \subset A - C$, and $B - A \subset C - A$. Furthermore, suppose s also satisfies the present theory and assume (for simplicity) that both d and s are symmetric. According to the representation theorem, therefore, there exist a nonnegative scale f and nonnegative constants θ and λ such that for all a,b,c,e,

$$s(a,b) > s(c,e) \quad \text{iff} \quad \theta f(A \cap B) - f(A - B) - f(B - A) > \theta f(C \cap E) - f(C - E) - f(E - C),$$

and

$$d(a,b) > d(c,e) \quad \text{iff} \quad f(A - B) + f(B - A) - \lambda f(A \cap B) > f(C - E) + f(E - C) - \lambda f(C \cap E).$$

The weights associated with the distinctive features can be set equal to 1 in the symmetric case with no loss of generality. Hence, θ and λ reflect the *relative* weight of the common features in the assessment of similarity and difference, respectively.

Note that if θ is very large then the similarity ordering is essentially determined by the common features. On the other hand, if λ is very small, then the difference ordering is determined primarily by the distinctive features. Consequently, both $s(a,b) > s(c,e)$ and $d(a,b) > d(c,e)$ may be obtained whenever

$$f(A \cap B) > f(C \cap E)$$

and

$$f(A - B) + f(B - A) > f(C - E) + f(E - C).$$

That is, if the common features are weighed more heavily in judgments of similarity than in judgments of difference, then a pair of objects with many common and many distinctive features may be perceived as both more similar and more different than another pair of objects with fewer common and fewer distinctive features.

To test this hypothesis, 20 sets of four countries were constructed on the basis of a pilot test. Each set included two pairs of countries: a prominent pair and a nonprominent pair. The prominent pairs consisted of countries that were well known to our subjects (e.g., USA–USSR, Red China–Japan). The nonprominent pairs consisted of countries that were known to the subjects, but not as well as the prominent ones (e.g., Tunis–Morocco, Paraguay–Ecuador). All subjects were presented with the same 20 sets. One group of 30 subjects selected between the two pairs in each set the pair of countries that were more *similar*. Another group of 30 subjects selected between the two pairs in each set the pair of countries that were more *different*.

Let Π_s and Π_d denote, respectively, the percentage of choices where the prominent pair of countries was selected as more similar or as more different. If similarity and difference are complementary (i.e., $\theta = \lambda$), then $\Pi_s + \Pi_d$ should equal 100 for all pairs. On the other hand, if $\theta > \lambda$, then $\Pi_s + \Pi_d$ should exceed 100. The average value of $\Pi_s + \Pi_d$, across all sets, was 113.5, which is significantly greater than 100, $t(59) = 3.27$, $p < .01$.

Moreover, on the average, the prominent pairs were selected more frequently than the nonprominent pairs in both the similarity and the difference tasks. For example, 67% of the subjects in the similarity group selected West Germany and East Germany as more similar to each other than Ceylon and Nepal, while 70% of the subjects in the difference group selected West Germany and East Germany as more different from each other than Ceylon and Nepal. These data demonstrate how the relative weight of the common and the distinctive features varies with the task and support the hypothesis that people attend more to the common features in judgments of similarity than in judgments of difference.

Similarity in Context

Like other judgments, similarity depends on context and frame of reference. Sometimes the relevant frame of reference is specified explicitly, as in the questions, "How similar are English and French with respect to sound?" "What is the similarity of a pear and an apple with respect to taste?" In general, however, the relevant feature space is not specified explicitly but rather inferred from the general context.

When subjects are asked to assess the similarity between the USA and the USSR, for instance, they usually assume that the relevant context is the set of countries and that the relevant frame of reference includes all political, geographical, and cultural features. The relative weights assigned to these features, of course, may differ for different people. With natural, integral stimuli such as countries, people, colors, and sounds, there is relatively little ambiguity regarding the relevant feature space. However, with artificial, separable stimuli, such as figures varying in color and shape, or lines varying in length and orientation, subjects sometimes experience difficulty in evaluating overall similarity and occasionally tend to evaluate similarity with respect to one factor or the other (Shepard, 1964) or change the relative weights of attributes with a change in context (Torgerson, 1965).

In the present theory, changes in context or frame of reference correspond to changes in the measure of the feature space. When asked to assess the political similarity between countries, for example, the subject presumably attends to the political aspects of the countries and ignores, or assigns a weight of zero to, all other features. In addition to such restrictions of the feature space induced by explicit or implicit instructions, the salience of features and hence the similarity of objects are also influenced by the effective context (i.e., the

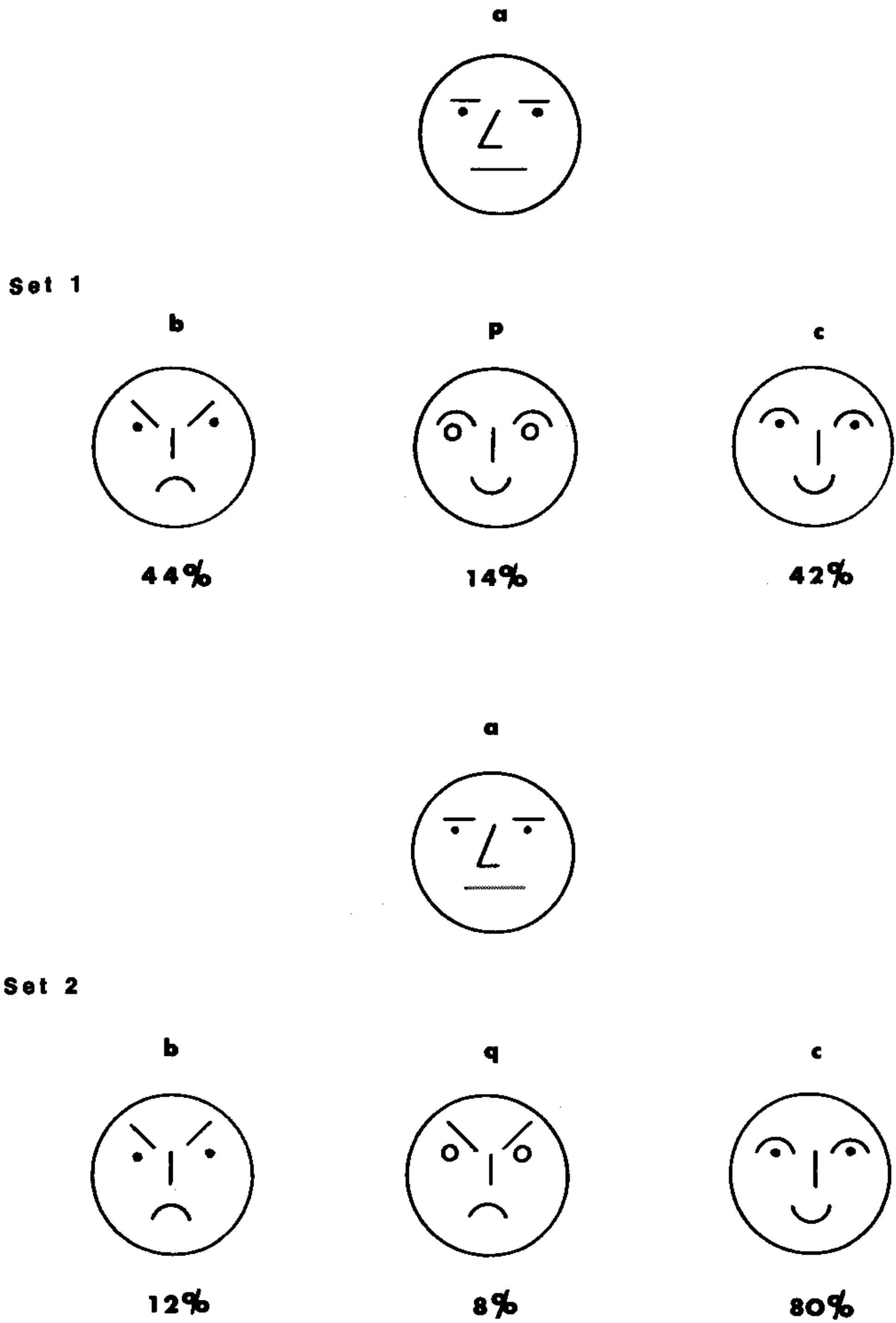


Figure 4. Two sets of schematic faces used to test the diagnosticity hypothesis. The percentage of subjects who selected each face (as most similar to the target) is presented below the face.

set of objects under consideration). To understand this process, let us examine the factors that determine the salience of a feature and its contribution to the similarity of objects.

The Diagnosticity Principle

The salience (or the measure) of a feature is determined by two types of factors: intensive and diagnostic. The former refers to factors that increase intensity or signal-to-noise ratio, such as the brightness of a light, the loudness of a tone, the saturation of a color, the size of a letter, the frequency of an item, the clarity of a picture, or the vividness of an image. The diagnostic factors refer to the classificatory significance of features, that is, the importance or prevalence of the classifications that are based on these features. Unlike the intensive factors, the diagnostic factors are highly sensitive to the particular object set under study. For example, the feature "real" has no diagnostic value in the set of actual animals since it is shared by all actual animals and hence cannot be used to classify them. This feature, however, acquires considerable diagnostic value if the object set is extended to include legendary animals, such as a centaur, a mermaid, or a phoenix.

When faced with a set of objects, people often sort them into clusters to reduce information load and facilitate further processing. Clusters are typically selected so as to maximize the similarity of objects within a cluster and the dissimilarity of objects from different clusters. Hence, the addition and/or deletion of objects can alter the clustering of the remaining objects. A change of clusters, in turn, is expected to increase the diagnostic value of features on which the new clusters are based, and therefore, the similarity of objects that share these features. This relation between similarity and grouping—called the diagnosticity hypothesis—is best explained in terms of a concrete example. Consider the two sets of four schematic faces (displayed in Figure 4), which differ in only one of their elements (p and q).

The four faces of each set were displayed in a row and presented to a different group of 25 subjects who were instructed to partition them into two pairs. The most frequent partition of Set 1 was c and p (smiling faces) versus a and b (nonsmiling faces). The most common partition of Set 2 was b and q (frowning faces) versus a and c (nonfrowning faces). Thus, the replacement of p by q changed the grouping of a: In Set 1 a was paired with b, while in Set 2 a was paired with c.

According to the above analysis, smiling has a greater diagnostic value in Set 1 than in Set 2, whereas frowning has a greater diagnostic value in Set 2 than in Set 1. By the diagnosticity hypothesis, therefore, similarity should follow the grouping. That is, the similarity of a (which has a neutral expression) to b (which is frowning) should be greater in Set 1, where they are grouped together, than in Set 2, where they are grouped separately. Likewise, the similarity of a to c (which is smiling) should be greater in Set 2, where they are grouped together, than in Set 1, where they are not.

To test this prediction, two different groups of 50 subjects were presented with Sets 1 and 2 (in the form displayed in Figure 4) and asked to select one of the three faces below (called the choice set) that was most similar to the face on the top (called the target). The percentage of subjects who selected each of the three elements of the choice set is presented below the face. The results confirmed the diagnosticity hypothesis: b was chosen more frequently in Set 1 than in Set 2, whereas c was chosen more frequently in Set 2 than in Set 1. Both differences are statistically significant, $p < .01$. Moreover, the replacement of p by q actually reversed the similarity ordering: In Set 1, b is more similar to a than c, whereas in Set 2, c is more similar to a than b.

A more extensive test of the diagnosticity hypothesis was conducted using semantic rather than visual stimuli. The experimental design was essentially the same, except that countries served as stimuli instead of faces. Twenty pairs of matched sets of four countries of the form {a,b,c,p} and {a,b,c,q} were constructed. An example of two matched sets is presented in Figure 5.

Note that the two matched sets (1 and 2) differ only by one element (p and q). The sets were constructed so that a (in this case Austria) is likely to be grouped with b (e.g., Sweden) in Set 1, and with c (e.g., Hungary) in Set 2. To validate this assumption, we presented two groups of 25 subjects with all sets of four countries and asked them to partition each quadruple into two pairs. Each group received one of the two matched quadruples, which were displayed in a row in random order. The results confirmed our prior hypothesis regarding the grouping of countries. In every case but one, the replacement of p by q changed the pairing of the target country in the predicted direction, $p < .01$ by sign test. For example, Austria was paired with Sweden by 60% of the subjects in Set 1, and it was paired with Hungary by 96% of the subjects in Set 2.

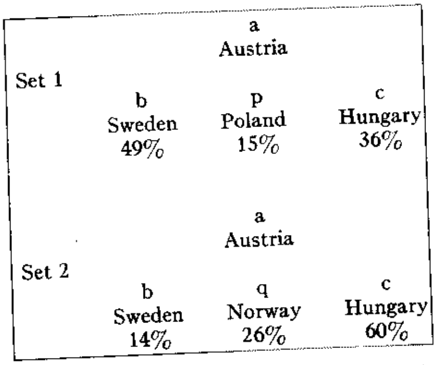


Figure 5. Two sets of countries used to test the diagnosticity hypothesis. The percentage of subjects who selected each country (as most similar to Austria) is presented below the country.

To test the diagnosticity hypothesis, we presented two groups of 35 subjects with 20 sets of four countries in the format displayed in Figure 5. These subjects were asked to select, for each quadruple, the country in the choice set that was most similar to the target country. Each group received exactly one quadruple from each pair. If the similarity of b to a, say, is independent of the choice set, then the proportion of subjects who chose b rather than c as most similar to a should be the same regardless of whether the third element in the choice set is p or q. For example, the proportion of subjects who select Sweden rather than Hungary as most similar to Austria should be independent of whether the odd element in the choice set is Norway or Poland.

In contrast, the diagnosticity hypothesis implies that the change in grouping, induced by the substitution of the odd element, will change the similarities in a predictable manner. Recall that in Set 1 Poland was paired with Hungary, and Austria with Sweden, while in Set 2 Norway was paired with Sweden, and Austria with Hungary. Hence, the proportion of subjects who select Sweden rather than Hungary (as most similar to Austria) should be higher in Set 1 than in Set 2. This prediction is strongly supported by the data in Figure 5, which show that Sweden was selected more frequently than Hungary in Set 1, while Hungary was selected more frequently than Sweden in Set 2.

Let b(p) denote the percentage of subjects who chose country b as most similar to a when the odd element in the choice set is p, and so on. As in the above examples, the notation is chosen so that b is generally grouped with q, and c is generally grouped with p. The differences $b(p) - b(q)$ and $c(q) - c(p)$, therefore, reflect the effects of the odd elements, p and q, on the similarity of b and c to the target a. In the absence of context effects, both differences should equal 0, while under the diagnosticity hypothesis both differences should be positive. In Figure 5, for example, $b(p) - b(q) = 49 - 14 = 35$, and $c(q) - c(p) = 60 - 36 = 24$. The average difference, across all pairs of quadruples, equals 9%, which is significantly positive, $t(19) = 3.65$, $p < .01$.

Several variations of the experiment did not alter the nature of the results, The diagnosticity hypothesis was also confirmed when (i) each choice set contained four elements, rather than three, (ii) the subjects were instructed to rank the elements of each choice set according to their similarity to the target, rather than to select the most similar element, and (iii) the target consisted of two elements, and the subjects were instructed to select one element of the choice set that was most similar to the two target elements. For further details, see Tversky and Gati (in press).

The Extension Effect

Recall that the diagnosticity of features is determined by the classifications that are based on them. Features that are shared by all the objects under consideration cannot be used to classify these objects and are, therefore, devoid of diagnostic value. When the context is extended by the enlargement of the object set, some features that had been shared by all objects in the original context may not be shared by all objects in the broader context. These features then acquire diagnostic value and increase the similarity of the objects that share them. Thus, the similarity of a pair of objects in the original context will usually be smaller than their similarity in the extended context.

Essentially the same account was proposed and supported by Sjöberg (Note 1) in studies of similarity between animals, and between musical instruments. For example, Sjöberg showed that the similarities between string instruments (banjo, violin, harp, electric guitar) were increased when a wind instrument (clarinet) was added to this set. Since the string instruments are more similar to each other than to the clarinet, however, the above result may be attributed, in part at least, to subjects' tendency to standardize the response scale, that is, to produce the same average similarity for any set of comparisons.

This effect can be eliminated by the use of a somewhat different design, employed in the following study. Subjects were presented with pairs of countries having a common border and assessed their similarity on a 20-point scale. Four sets of eight pairs were constructed. Set 1 contained eight pairs of European countries (e.g., Italy–Switzerland). Set 2 contained eight pairs of American countries (e.g., Brazil–Uruguay). Set 3 contained four pairs from Set 1 and four pairs from Set 2, while Set 4 contained the remaining pairs from Sets 1 and 2. Each one of the four sets was presented to a different group of 30–36 subjects.

According to the diagnosticity hypothesis, the features "European" and "American" have no diagnostic value in Sets 1 and 2, although they both have a diagnostic value in Sets 3 and 4. Consequently, the overall average similarity in the heterogeneous sets (3 and 4) is expected to be higher than the overall average similarity in the homogeneous sets (1 and 2). This prediction was confirmed by the data, $t(15) = 2.11$, $p < .05$.

In the present study all similarity assessments involve only homogeneous pairs (i.e., pairs of countries from the same continent sharing a common border). Unlike Sjöberg's (Note 1) study, which extended the context by introducing nonhomogeneous pairs, our experiment extended the context by constructing heterogeneous sets composed of homogeneous pairs. Hence, the increase of similarity with the enlargement of context, observed in the present study, cannot be explained by subjects' tendency to equate the average similarity for any set of assessments.

The Two Faces of Similarity

According to the present analysis, the salience of features has two components: intensity and diagnosticity. The intensity of a feature is determined by perceptual and cognitive factors that are relatively stable across contexts. The diagnostic value of a feature is determined by the prevalence of the classifications that are based on it, which change with the context. The effects of context on similarity, therefore, are treated as changes in the diagnostic value of features induced by the respective changes in the grouping of the objects.

This account was supported by the experimental finding that changes in grouping (produced by the replacement or addition of objects) lead to corresponding changes in the similarity of the objects. These results shed light on the dynamic interplay between similarity and classification. It is generally assumed that classifications are determined by similarities among the objects. The preceding discussion supports the converse hypothesis: that the similarity of objects is modified by the manner in which they are classified. Thus, similarity has two faces: causal and derivative. It serves as a basis for the classification of objects, but it is also influenced by the adopted classification. The diagnosticity principle which underlies this process may provide a key to the analysis of the effects of context on similarity.

Discussion

In this section we relate the present development to the representation of objects in terms of clusters and trees, discuss the con-

Table 1
ADCLUS Analysis of the Similarities Among the Integers 0 Through 9 (from Shepard & Arabie, Note 2)

Rank	Weight	Elements of subset	Interpretation of subset
1st	.305	2 4 8	powers of two
2nd	.288	6 7 8 9	large numbers
3rd	.279	3 6 9	multiples of three
4th	.202	0 1 2	very small numbers
5th	.202	1 3 5 7 9	odd numbers
6th	.175	1 2 3	small nonzero numbers
7th	.163	5 6 7	middle numbers (largish)
8th	.160	0 1	additive and multiplicative identities
9th	.146	0 1 2 3 4	smallish numbers

cepts of prototypicality and family resemblance, and comment on the relation between similarity and metaphor.

Features, Clusters, and Trees

There is a well-known correspondence between features or properties of objects and the classes to which the objects belong. A red flower, for example, can be characterized as having the feature "red," or as being a member of the class of red objects. In this manner we associate with every feature in Φ the class of objects in Δ which possesses that feature. This correspondence between features and classes provides a direct link between the present theory and the clustering approach to the representation of proximity data.

In the contrast model, the similarity between objects is expressed as a function of their common and distinctive features. Relations among overlapping sets are often represented in a Venn diagram (see Figure 1). However, this representation becomes cumbersome when the number of objects exceeds four or five. To obtain useful graphic representations of the contrast model, two alternative simplifications are entertained.

First, suppose the objects under study are all equal in prominence, that is, $f(A) = f(B)$ for all a,b in Δ. Although this assumption is not strictly valid in general, it may serve as a reasonable approximation in certain contexts. Assuming feature additivity and symmetry, we obtain

$$\begin{aligned} S(a,b) &= \theta f(A \cap B) - f(A - B) - f(B - A) \\ &= \theta f(A \cap B) + 2f(A \cap B) - f(A - B) \\ &\quad - f(B - A) - 2f(A \cap B) \\ &= (\theta + 2)f(A \cap B) - f(A) - f(B) \\ &= \lambda f(A \cap B) + \mu, \end{aligned}$$

since $f(A) = f(B)$ for all a,b in Δ. Under the present assumptions, therefore, similarity between objects is a linear function of the measure of their common features.

Since f is an additive measure, $f(A \cap B)$ is expressible as the sum of the measures of all the features that belong to both a and b. For each subset Λ of Δ, let $\Phi(\Lambda)$ denote the set of features that are shared by all objects in Λ, and are not shared by any object that does not belong to Λ. Hence,

$$\begin{aligned} S(a,b) &= \lambda f(A \cap B) + \mu \\ &= \lambda\Big(\sum_{X \in A \cap B} f(X)\Big) + \mu \\ &= \lambda\Big(\sum_{\Lambda \supset \{a,b\}} f(\Phi(\Lambda))\Big) + \mu . \end{aligned}$$

Since the summation ranges over all subsets of Δ that include both a and b, the similarity between objects can be expressed as the sum of the weights associated with all the sets that include both objects.

This form is essentially identical to the additive clustering model proposed by Shepard and Arabie (Note 2). These investigators have developed a computer program, ADCLUS, which selects a relatively small collection of subsets

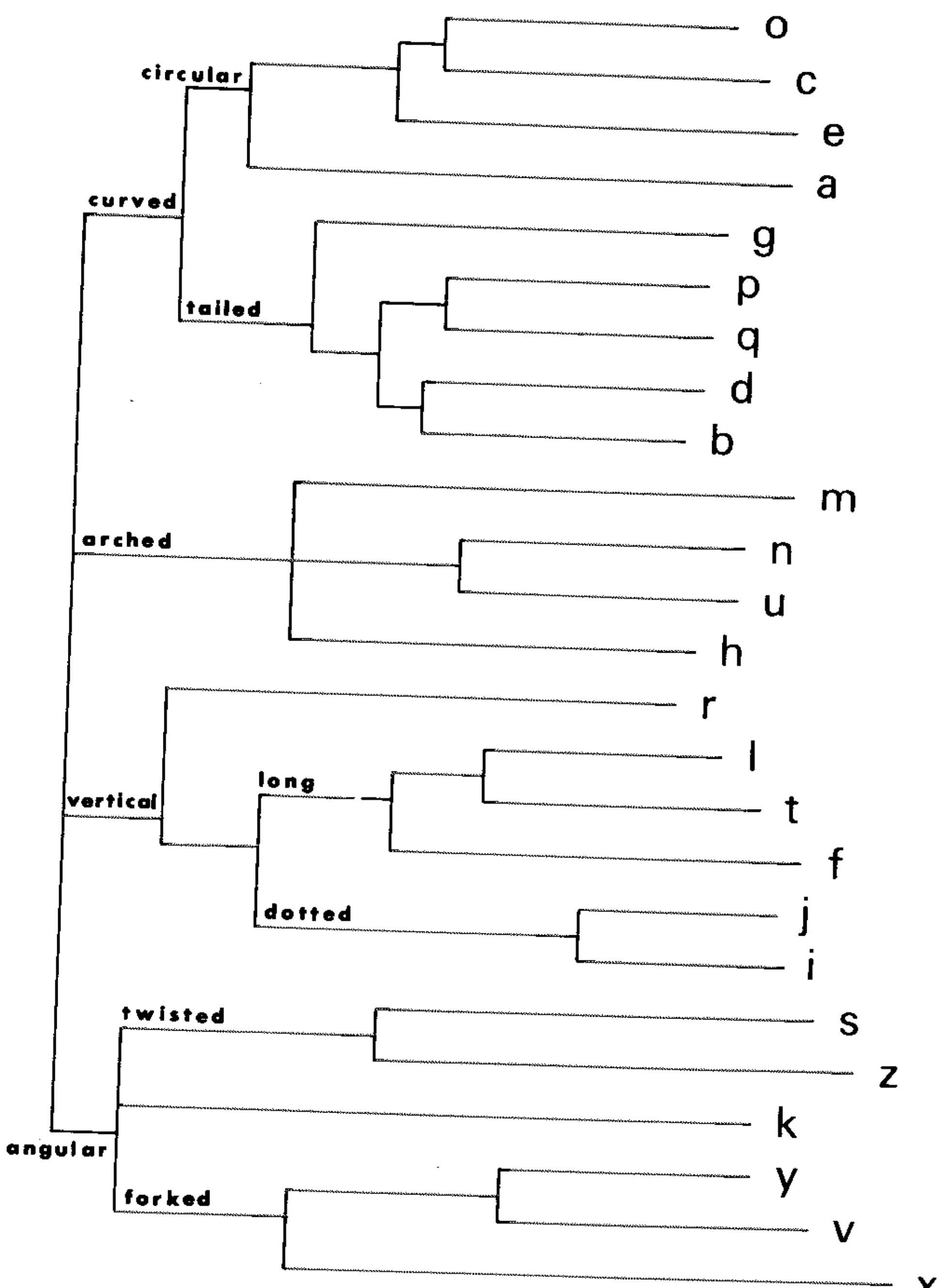


Figure 6. The representation of letter similarity as an additive (feature) tree. From Sattath and Tversky (in press).

and assigns weight to each subset so as to maximize the proportion of (similarity) variance accounted for by the model. Shepard and Arabie (Note 2) applied ADCLUS to several studies including Shepard, Kilpatric, and Cunningham's (1975) on judgments of similarity between the integers 0 through 9 with respect to their abstract numerical character.

A solution with 19 subsets accounted for 95% of the variance. The nine major subsets (with the largest weights) are displayed in Table 1 along with a suggested interpretation. Note that all the major subsets are readily interpretable, and they are overlapping rather than hierarchical.

The above model expresses similarity in terms of common features only. Alternatively, similarity may be expressed exclusively in terms of distinctive features. It has been shown by Sattath (Note 3) that for any symmetric contrast model with an additive measure f, there exists a measure g defined on the same feature space such that

$$\begin{aligned} S(a,b) &= \theta f(A \cap B) - f(A - B) - f(B - A) \\ &= \lambda - g(A - B) - g(B - A) \end{aligned}$$
$$\text{for some } \lambda > 0.$$

This result allows a simple representation of dissimilarity whenever the feature space Φ is a tree (i.e., whenever any three objects in Δ can be labeled so that $A \cap B = A \cap C \subset B \cap C$). Figure 6 presents an example of a feature tree, constructed by Sattath and Tversky (in press) from judged similarities between lowercase letters, obtained by Kuennapas and Janson (1969). The major branches are labeled to facilitate the interpretation of the tree.

Each (horizontal) arc in the graph represents the set of features shared by all the objects (i.e., letters) that follow from that arc, and the arc length corresponds to the measure of that set. The features of an object are the features of all the arcs which lead to that object, and its measure is its (horizontal) distance to the root. The tree distance between objects a and b is the (horizontal) length of the path joining them, that is, $f(A - B) + f(B - A)$. Hence, if the contrast model holds, $\alpha = \beta$, and Φ is a tree, then dissimilarity (i.e., $-S$) is expressible as tree distance.

A feature tree can also be interpreted as a hierarchical clustering scheme where each arc length represents the weight of the cluster consisting of all the objects that follow from that arc. Note that the tree in Figure 6 differs from the common hierarchical clustering tree in that the branches differ in length. Sattath and Tversky (in press) describe a computer program, ADDTREE, for the construction of additive feature trees from similarity data and discuss its relation to other scaling methods.

It follows readily from the above discussion that if we assume both that the feature set Φ is a tree, and that $f(A) = f(B)$ for all a,b in Δ, then the contrast model reduces to the well-known hierarchical clustering scheme. Hence, the additive clustering model (Shepard & Arabie, Note 2), the additive similarity tree (Sattath & Tversky, in press), and the hierarchical clustering scheme (Johnson, 1967) are all special cases of the contrast model. These scaling models can thus be used to discover the common and distinctive features of the objects under study. The present development, in turn, provides theoretical foundations for the analysis of set-theoretical methods for the representation of proximities.

Similarity, Prototypicality, and Family Resemblance

Similarity is a relation of proximity that holds between two objects. There exist other proximity relations such as prototypicality and representativeness that hold between an object and a class. Intuitively, an object is prototypical if it exemplifies the category to which it belongs. Note that the prototype is not necessarily the most typical or frequent member of its class. Recent research has demonstrated the importance of prototypicality or representativeness in perceptual learning (Posner & Keele, 1968; Reed, 1972), inductive inference (Kahneman & Tversky, 1973), semantic memory (Smith, Rips, & Shoben, 1974), and the formation of categories (Rosch & Mervis, 1975). The following discussion analyzes the relations of prototypicality and family resemblance in terms of the present theory of similarity.

Let $P(a,\Lambda)$ denote the (degree of) prototypicality of object a with respect to class Λ, with cardinality n, defined by

$$P(a,\Lambda) = p_n(\lambda\Sigma f(A \cap B) - \Sigma(f(A - B) + f(B - A))),$$

where the summations are over all b in Λ. Thus, $P(a,\Lambda)$ is defined as a linear combination (i.e., a contrast) of the measures of the features of a that are shared with the elements of Λ and the features of a that are not shared with the elements of Λ. An element a of Λ is a *prototype* if it maximizes P(a,Λ). Note that a class may have more than one prototype.

The factor p_n reflects the effect of category size on prototypicality, and the constant λ determines the relative weights of the common and the distinctive features. If $p_n = 1/n$, $\lambda = \theta$, and $\alpha = \beta = 1$, then $P(a,\Lambda) = 1/n\Sigma S(a,b)$ (i.e., the prototypicality of a with respect to Λ equals the average similarity of a to all members of Λ). However, in line with the focusing hypotheses discussed earlier, it appears likely that the common features are weighted more heavily in judgments of prototypicality than in judgments of similarity.

Some evidence concerning the validity of the proposed measure was reported by Rosch and Mervis (1975). They selected 20 objects from each one of six categories (furniture, vehicle, fruit, weapon, vegetable, clothing) and instructed subjects to list the attributes associated with each one of the objects. The prototypicality of an object was defined by the number of attributes or features it shared with each member of the category. Hence, the prototypicality of a with respect to Λ was defined by ΣN(a,b), where N(a,b) denotes the number of attributes shared by a and b, and the summation ranges over all b in Λ. Clearly, the measure of prototypicality employed by Rosch and Mervis (1975) is a special case of the proposed measure, where λ is large and $f(A \cap B) = N(a,b)$.

These investigators also obtained direct measures of prototypicality by instructing subjects to rate each object on a 7-point scale according to the extent to which it fits the "idea or image of the meaning of the category." The rank correlations between these ratings and the above measure were quite high in all categories: furniture, .88; vehicle, .92; weapon, .94; fruit, .85; vegetable, .84; clothing, .91. The rated prototypicality of an object in a category, therefore, is predictable by the number of features it shares with other members of that category.

In contrast to the view that natural categories are definable by a conjunction of critical features, Wittgenstein (1953) argued that several natural categories (e.g., a game) do not have any attribute that is shared by all their members, and by them alone. Wittgenstein proposed that natural categories and concepts are commonly characterized and understood in terms of family resemblance, that is, a network of similarity relations that link the various members of the class. The importance of family resemblance in the formation and processing of categories has been effectively underscored by the work of Rosch and her collaborators (Rosch, 1973; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). This research demonstrated that both natural and artificial categories are commonly perceived and organized in terms of prototypes, or focal elements, and some measure of proximity from the prototypes. Furthermore, it lent substantial support to the claim that people structure their world in terms of basic semantic categories that represent an optimal level of abstraction. Chair, for example, is a basic category; furniture is too general and kitchen chair is too specific. Similarly, car is a basic category; vehicle is too general and sedan is too specific. Rosch argued that the basic categories are selected so as to maximize family resemblance—defined in terms of cue validity.

The present development suggests the following measure for family resemblance, or category resemblance. Let Λ be some subset of Δ with cardinality n. The category resemblance of Λ denoted R(Λ) is defined by

$$R(\Lambda) = r_n(\lambda\Sigma f(A \cap B) - \Sigma(f(A - B) + f(B - A))),$$

the summations being over all a,b in Λ. Hence, category resemblance is a linear combination of the measures of the common and the distinctive features of all pairs of objects in that category. The factor r_n reflects the effect of category size on category resemblance, and the constant λ determines the *relative* weight of the common and the distinctive features. If $\lambda = \theta$, $\alpha = \beta = 1$, and $r_n = 2/n(n - 1)$, then

$$R(\Lambda) = \frac{\Sigma S(a,b)}{\binom{n}{2}},$$

the summation being over all a,b in Λ; that is, category resemblance equals average similarity between all members of Λ. Although the proposed measure of family resemblance differs from Rosch's, it nevertheless captures her basic notion that family resemblance is highest for those categories which "have the most attributes common to members of the category and the least attributes shared with members of other categories" (Rosch et al., 1976, p. 435).

The maximization of category resemblance could be used to explain the formation of categories. Thus, the set Λ rather than Γ is selected as a natural category whenever $R(\Lambda) > R(\Gamma)$. Equivalently, an object a is added to a category Λ whenever $R(\{\Lambda \cup a\}) > R(\Lambda)$. The fact that the preferred (basic) categories are neither the most inclusive nor the most specific imposes certain constraints on r_n.

If $r_n = 2/n(n-1)$ then $R(\Lambda)$ equals the average similarity between all members of Λ. This index leads to the selection of minimal categories because average similarity can generally be increased by deleting elements. The average similarity between sedans, for example, is surely greater than the average similarity between cars; nevertheless, car rather than sedan serves as a basic category. If $r_n = 1$ then $R(\Lambda)$ equals the sum of the similarities between all members of Λ. This index leads to the selection of maximal categories because the addition of objects increases total similarity, provided S is nonnegative.

In order to explain the formation of intermediate-level categories, therefore, category resemblance must be a compromise between an average and a sum. That is, r_n must be a decreasing function of n that exceeds $2/n(n-1)$. In this case, $R(\Lambda)$ increases with category size whenever average similarity is held constant, and vice versa. Thus, a considerable increase in the extension of a category could outweigh a small reduction in average similarity.

Although the concepts of similarity, prototypicality, and family resemblance are intimately connected, they have not been previously related in a formal explicit manner. The present development offers explications of similarity, prototypicality, and family resemblance within a unified framework, in which they are viewed as contrasts, or linear combinations, of the measures of the appropriate sets of common and distinctive features.

Similes and Metaphors

Similes and metaphors are essential ingredients of creative verbal expression. Perhaps the most interesting property of metaphoric expressions is that despite their novelty and nonliteral nature, they are usually understandable and often informative. For example, the statement that Mr. X resembles a bulldozer is readily understood as saying that Mr. X is a gross, powerful person who overcomes all obstacles in getting a job done. An adequate analysis of connotative meaning should account for man's ability to interpret metaphors without specific prior learning. Since the message conveyed by such expressions is often pointed and specific, they cannot be explained in terms of a few generalized dimensions of connotative meaning, such as evaluation or potency (Osgood, 1962). It appears that people interpret similes by scanning the feature space and selecting the features of the referent that are applicable to the subject (e.g., by selecting features of the bulldozer that are applicable to the person). The nature of this process is left to be explained.

There is a close tie between the assessment of similarity and the interpretation of metaphors. In judgments of similarity one assumes a particular feature space, or a frame of reference, and assesses the quality of the match between the subject and the referent. In the interpretation of similes, one assumes a resemblance between the subject and the referent and searches for an interpretation of the space that would maximize the quality of the match. The same pair of objects, therefore, can be viewed as similar or different depending on the choice of a frame of reference.

One characteristic of good metaphors is the contrast between the prior, literal interpretation, and the posterior, metaphoric interpretation. Metaphors that are too transparent are uninteresting; obscure metaphors are uninterpretable. A good metaphor is like a good detective story. The solution should not be apparent in advance to maintain the reader's interest, yet it should seem plausible after the fact to maintain coherence of the story. Consider the simile "An essay is like a fish." At first, the statement is puzzling. An essay is not expected to be fishy, slippery, or wet. The puzzle is resolved when we recall that (like a fish) an essay has a head and a body, and it occasionally ends with a flip of the tail.

Appendix

An Axiomatic Theory of Similarity

Let $\Delta = \{a,b,c,\ldots\}$ be a collection of objects characterized as sets of features, and let A,B,C, denote the sets of features associated with a,b,c, respectively. Let s(a,b) be an ordinal measure of the similarity of a to b, defined for all distinct a,b in Δ. The present theory is based on the following five axioms. Since the first three axioms are discussed in the paper, they are merely restated here; the remaining axioms are briefly discussed.

1. *Matching*: $s(a,b) = F(A \cap B,\ A - B,\ B - A)$ where F is some real-valued function in three arguments.

2. *Monotonicity*: $s(a,b) \geq s(a,c)$ whenever $A \cap B \supset A \cap C$, $A - B \subset A - C$, and $B - A \subset C - A$. Moreover, if either inclusion is proper then the inequality is strict.

Let Φ be the set of all features associated with the objects of Δ, and let X,Y,Z, etc. denote subsets of Φ. The expression F(X,Y,Z) is defined whenever there exist a,b in Δ such that $A \cap B = X$, $A - B = Y$, and $B - A = Z$, whence $s(a,b) = F(X,Y,Z)$. Define $V \simeq W$ if one or more of the following hold for some X,Y,Z: $F(V,Y,Z) = F(W,Y,Z)$, $F(X,V,Z) = F(X,W,Z)$, $F(X,Y,V) = F(X,Y,W)$. The pairs (a,b) and (c,d) *agree* on one, two, or three components, respectively, whenever one, two, or three of the following hold: $(A \cap B) \simeq (C \cap D)$, $(A - B) \simeq (C - D)$, $(B - A) \simeq (D - C)$.

3. *Independence*: Suppose the pairs (a,b) and (c,d), as well as the pairs (a′,b′) and (c′,d′), agree on the same two components, while the pairs (a,b) and (a′,b′), as well as the pairs (c,d) and (c′,d′), agree on the remaining (third) component. Then

$$s(a,b) \geq s(a',b') \quad \text{iff} \quad s(c,d) \geq s(c',d').$$

4. *Solvability*:

(i). For all pairs (a,b), (c,d), (e,f), of objects in Δ there exists a pair (p,q) which agrees with them, respectively, on the first, second, and third component, that is, $P \cap Q \simeq A \cap B$, $P - Q \simeq C - D$, and $Q - P \simeq F - E$.

(ii). Suppose $s(a,b) > t > s(c,d)$. Then there exist e,f with $s(e,f) = t$, such that if (a,b) and (c,d) agree on one or two components, then (e,f) agrees with them on these components.

(iii). There exist pairs (a,b) and (c,d) of objects in Δ that do not agree on any component.

Unlike the other axioms, solvability does not impose constraints on the similarity order; it merely asserts that the structure under study is sufficiently rich so that certain equations can be solved. The first part of Axiom 4 is analogous to the existence of a factorial structure. The second part of the axiom implies that the range of s is a real interval: There exist objects in Δ whose similarity matches any real value that is bounded by two similarities. The third part of Axiom 4 ensures that all arguments of F are essential.

Let Φ_1, Φ_2, and Φ_3 be the sets of features that appear, respectively, as first, second, or third arguments of F. (Note that $\Phi_2 = \Phi_3$.) Suppose X and X′ belong to Φ_1, while Y and Y′ belong to Φ_2. Define $(X,X')_1 \simeq (Y,Y')_2$ whenever the two intervals are matched, that is, whenever there exist pairs (a,b) and (a′,b′) of equally similar objects in Δ which agree on the third factor. Thus, $(X,X')_1 \simeq (Y,Y')_2$ whenever

$$s(a,b) = F(X,Y,Z) = F(X',Y',Z) = s(a',b').$$

This definition is readily extended to any other pair of factors. Next, define $(V,V')_i \simeq (W,W')_i$, $i = 1,2,3$ whenever $(V,V')_i \simeq (X,X')_j \simeq (W,W')_i$, for some $(X,X')_j$, $j \neq i$. Thus, two intervals on the same factor are equivalent if both match the same interval on another factor. The following invariance axiom asserts that if two intervals are equivalent on one factor, they are also equivalent on another factor.

5. *Invariance*: Suppose V,V′, W,W′ belong to both Φ_i and Φ_j, i,j = 1,2,3. Then

$$(V,V')_i \simeq (W,W')_i \quad \text{iff} \quad (V,V')_j \simeq (W,W')_j.$$

Representation Theorem

Suppose Axioms 1–5 hold. Then there exist a similarity scale S and a nonnegative scale f such that for all a,b,c,d in Δ

(i). $S(a,b) \geq S(c,d)$ iff $s(a,b) \geq s(c,d)$,
(ii). $S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$, for some $\theta,\alpha,\beta \geq 0$.
(iii). f and S are interval scales.

While a self-contained proof of the representation theorem is quite long, the theorem can be readily reduced to previous results.

Recall that Φ_i is the set of features that appear as the *i*th argument of F, and let $\Psi_i = \Phi_i/\simeq$, i = 1,2,3. Thus, Ψ_i is the set of equivalence classes of Φ_i with respect to $\simeq$. It follows from Axioms 1 and 3 that each Ψ_i is well defined, and it follows from Axiom 4 that $\Psi = \Psi_1 \times \Psi_2$

$\times \Psi_3$ is equivalent to the domain of F. We wish to show that Ψ, ordered by F, is a three-component, additive conjoint structure, in the sense of Krantz, Luce, Suppes, and Tversky (1971, Section 6.11.1).

This result, however, follows from the analysis of decomposable similarity structures, developed by Tversky and Krantz (1970). In particular, the proof of part (c) of Theorem 1 in that paper implies that, under Axioms 1, 3, and 4, there exist nonnegative functions f_i defined on Ψ_i, $i = 1,2,3$, so that for all a,b,c,d in Δ

$$s(a,b) \geq s(c,d) \quad \text{iff} \quad S(a,b) \geq S(c,d)$$

where $S(a,b) = f_1(A \cap B) + f_2(A - B) + f_3(B - A)$,

and f_1, f_2, f_3 are interval scales with a common unit.

According to Axiom 5, the equivalence of intervals is preserved across factors. That is, fo all V,V', W,W' in $\Phi_i \cap \Phi_j$, $i,j, = 1,2,3$,

$$f_i(V) - f_i(V') = f_i(W) - f_i(W') \quad \text{iff} \quad f_j(V) - f_j(V') = f_j(W) - f_j(W')$$

Hence by part (i) of Theorem 6.15 of Krant et al. (1971), there exist a scale f and constant θ_i such that $f_i(X) = \theta_i f(X)$, $i = 1,2,3$. Finally by Axiom 2, S increases in f_1 and decreases in f_2 and f_3. Hence, it is expressible as

$$S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

for some nonnegative constants θ,α,β.

Received November 17, 1976

Amos Tversky
Stanford University

Daniel Kahneman
University of British Columbia

Rational Choice and the Framing of Decisions*

Alternative descriptions of a decision problem often give rise to different preferences, contrary to the principle of invariance that underlies the rational theory of choice. Violations of this theory are traced to the rules that govern the framing of decision and to the psychophysical principles of evaluation embodied in prospect theory. Invariance and dominance are obeyed when their application is transparent and often violated in other situations. Because these rules are normatively essential but descriptively invalid, no theory of choice can be both normatively adequate and descriptively accurate.

The modern theory of decision making under risk emerged from a logical analysis of games of chance rather than from a psychological analysis of risk and value. The theory was conceived as a normative model of an idealized decision maker, not as a description of the behavior of real people. In Schumpeter's words, it "has a much better claim to being called a logic of choice than a psychology of value" (1954, p. 1058).

The use of a normative analysis to predict and explain actual behavior is defended by several arguments. First, people are generally thought to be effective in pursuing their goals, particularly when they have incentives and opportunities to learn from experience. It seems reasonable, then, to describe choice as a maximization process. Second, competition favors rational individuals and organizations. Optimal decisions increase the chances of survival in a competitive environment, and a minority of rational individuals can sometimes impose rationality on the

* This work was supported by contract N00014-84-K-0615 from the Office of Naval Research to Stanford University. The present article reviews our work on decision making under risk from a new perspective, discussed primarily in the first and last sections. Most of the empirical demonstrations have been reported in earlier publications. Problems 3, 4, 7, 8, and 12 are published here for the first time. Requests for reprints should be addressed to Amos Tversky, Department of Psychology, Stanford University, Stanford, California 94705, or to Daniel Kahneman, Department of Psychology, University of California, Berkeley, California 94720.

(*Journal of Business*, 1986, vol. 59, no. 4, pt. 2)

0021-9398/86/5904-0010$01.50

whole market. Third, the intuitive appeal of the axioms of rational choice makes it plausible that the theory derived from these axioms should provide an acceptable account of choice behavior.

The thesis of the present article is that, in spite of these a priori arguments, the logic of choice does not provide an adequate foundation for a descriptive theory of decision making. We argue that the deviations of actual behavior from the normative model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the normative system. We first sketch an analysis of the foundations of the theory of rational choice and then show that the most basic rules of the theory are commonly violated by decision makers. We conclude from these findings that the normative and the descriptive analyses cannot be reconciled. A descriptive model of choice is presented, which accounts for preferences that are anomalous in the normative theory.

I. A Hierarchy of Normative Rules

The major achievement of the modern theory of decision under risk is the derivation of the expected utility rule from simple principles of rational choice that make no reference to long-run considerations (von Neumann and Morgenstern 1944). The axiomatic analysis of the foundations of expected utility theory reveals four substantive assumptions—cancellation, transitivity, dominance, and invariance—besides the more technical assumptions of comparability and continuity. The substantive assumptions can be ordered by their normative appeal, from the cancellation condition, which has been challenged by many theorists, to invariance, which has been accepted by all. We briefly discuss these assumptions.

Cancellation. The key qualitative property that gives rise to expected utility theory is the "cancellation" or elimination of any state of the world that yields the same outcome regardless of one's choice. This notion has been captured by different formal properties, such as the substitution axiom of von Neumann and Morgenstern (1944), the extended sure-thing principle of Savage (1954), and the independence condition of Luce and Krantz (1971). Thus, if A is preferred to B, then the prospect of winning A if it rains tomorrow (and nothing otherwise) should be preferred to the prospect of winning B if it rains tomorrow because the two prospects yield the same outcome (nothing) if there is no rain tomorrow. Cancellation is necessary to represent preference between prospects as the maximization of expected utility. The main argument for cancellation is that only one state will actually be realized, which makes it reasonable to evaluate the outcomes of options separately for each state. The choice between options should therefore depend only on states in which they yield different outcomes.

Transitivity. A basic assumption in models of both risky and riskless choice is the transitivity of preference. This assumption is necessary and essentially sufficient for the representation of preference by an ordinal utility scale u such that A is preferred to B whenever $u(A) > u(B)$. Thus transitivity is satisfied if it is possible to assign to each option a value that does not depend on the other available options. Transitivity is likely to hold when the options are evaluated separately but not when the consequences of an option depend on the alternative to which it is compared, as implied, for example, by considerations of regret. A common argument for transitivity is that cyclic preferences can support a "money pump," in which the intransitive person is induced to pay for a series of exchanges that returns to the initial option.

Dominance. This is perhaps the most obvious principle of rational choice: if one option is better than another in one state and at least as good in all other states, the dominant option should be chosen. A slightly stronger condition—called stochastic dominance—asserts that, for unidimensional risky prospects, A is preferred to B if the cumulative distribution of A is to the right of the cumulative distribution of B. Dominance is both simpler and more compelling than cancellation and transitivity, and it serves as the cornerstone of the normative theory of choice.

Invariance. An essential condition for a theory of choice that claims normative status is the principle of invariance: different representations of the same choice problem should yield the same preference. That is, the preference between options should be independent of their description. Two characterizations that the decision maker, on reflection, would view as alternative descriptions of the same problem should lead to the same choice—even without the benefit of such reflection. This principle of invariance (or extensionality [Arrow 1982]), is so basic that it is tacitly assumed in the characterization of options rather than explicitly stated as a testable axiom. For example, decision models that describe the objects of choice as random variables all assume that alternative representations of the same random variables should be treated alike. Invariance captures the normative intuition that variations of form that do not affect the actual outcomes should not affect the choice. A related concept, called consequentialism, has been discussed by Hammond (1985).

The four principles underlying expected utility theory can be ordered by their normative appeal. Invariance and dominance seem essential, transitivity could be questioned, and cancellation has been rejected by many authors. Indeed, the ingenious counterexamples of Allais (1953) and Ellsberg (1961) led several theorists to abandon cancellation and the expectation principle in favor of more general representations. Most of these models assume transitivity, dominance, and invariance (e.g., Hansson 1975; Allais 1979; Hagen 1979; Machina 1982; Quiggin 1982; Weber 1982; Chew 1983; Fishburn 1983; Schmeidler 1984; Segal 1984; Yaari 1984; Luce and Narens 1985). Other developments abandon transitivity but maintain invariance and dominance (e.g., Bell 1982; Fishburn 1982, 1984; Loomes and Sugden 1982). These theorists responded to observed violations of cancellation and transitivity by weakening the normative theory in order to retain its status as a descriptive model. However, this strategy cannot be extended to the failures of dominance and invariance that we shall document. Because invariance and dominance are normatively essential and descriptively invalid, a theory of rational decision cannot provide an adequate description of choice behavior.

We next illustrate failures of invariance and dominance and then review a descriptive analysis that traces these failures to the joint effects of the rules that govern the framing of prospects, the evaluation of outcomes, and the weighting of probabilities. Several phenomena of choice that support the present account are described.

II. Failures of Invariance

In this section we consider two illustrative examples in which the condition of invariance is violated and discuss some of the factors that produce these violations.

The first example comes from a study of preferences between medical treatments (McNeil et al. 1982). Respondents were given statistical information about the outcomes of two treatments of lung cancer. The same statistics were presented to some respondents in terms of mortality rates and to others in terms of survival rates. The respondents then indicated their preferred treatment. The information was presented as follows.[1]

Problem 1 (Survival frame)

Surgery: Of 100 people having surgery 90 live through the post-operative period, 68 are alive at the end of the first year and 34 are alive at the end of five years.

Radiation Therapy: Of 100 people having radiation therapy all live through the treatment, 77 are alive at the end of one year and 22 are alive at the end of five years.

Problem 1 (Mortality frame)

Surgery: Of 100 people having surgery 10 die during surgery or the post-operative period, 32 die by the end of the first year and 66 die by the end of five years.

1. All problems are presented in the text exactly as they were presented to the participants in the experiments.

> Radiation Therapy: Of 100 people having radiation therapy, none die during treatment, 23 die by the end of one year and 78 die by the end of five years.

The inconsequential difference in formulation produced a marked effect. The overall percentage of respondents who favored radiation therapy rose from 18% in the survival frame ($N = 247$) to 44% in the mortality frame ($N = 336$). The advantage of radiation therapy over surgery evidently looms larger when stated as a reduction of the risk of immediate death from 10% to 0% rather than as an increase from 90% to 100% in the rate of survival. The framing effect was not smaller for experienced physicians or for statistically sophisticated business students than for a group of clinic patients.

Our next example concerns decisions between conjunctions of risky prospects with monetary outcomes. Each respondent made two choices, one between favorable prospects and one between unfavorable prospects (Tversky and Kahneman 1981, p. 454). It was assumed that the two selected prospects would be played independently.

> Problem 2 ($N = 150$). Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.
>
> Decision (i) Choose between:
> A. a sure gain of $240 [84%]
> B. 25% chance to gain $1000 and 75% chance to gain nothing [16%]
>
> Decision (ii) Choose between:
> C. a sure loss of $750 [13%]
> D. 75% chance to lose $1000 and 25% chance to lose nothing [87%]

The total number of respondents is denoted by N, and the percentage who chose each option is indicated in brackets. (Unless otherwise specified, the data were obtained from undergraduate students at Stanford University and at the University of British Columbia.) The majority choice in decision i is risk averse, while the majority choice in decision ii is risk seeking. This is a common pattern: choices involving gains are usually risk averse, and choices involving losses are often risk seeking—except when the probability of winning or losing is small (Fishburn and Kochenberger 1979; Kahneman and Tversky 1979; Hershey and Schoemaker 1980).

Because the subjects considered the two decisions simultaneously, they expressed, in effect, a preference for the portfolio A and D over the portfolio B and C. However, the preferred portfolio is actually dominated by the rejected one! The combined options are as follows.

> A & D: 25% chance to win $240 and 75% chance to lose $760.
> B & C: 25% chance to win $250 and 75% chance to lose $750.

When the options are presented in this aggregated form, the dominant option is invariably chosen. In the format of problem 2, however, 73% of respondents chose the dominated combination A and D, and only 3% chose B and C. The contrast between the two formats illustrates a violation of invariance. The findings also support the general point that failures of invariance are likely to produce violations of stochastic dominance and vice versa.

The respondents evidently evaluated decisions i and ii separately in problem 2, where they exhibited the standard pattern of risk aversion in gains and risk seeking in losses. People who are given these problems are very surprised to learn that the combination of two preferences that they considered quite reasonable led them to select a dominated option. The same pattern of results was also observed in a scaled-down version of problem 2, with real monetary payoff (see Tversky and Kahneman 1981, p. 458).

As illustrated by the preceding examples, variations in the framing of decision problems produce systematic violations of invariance and dominance that cannot be defended on normative grounds. ·It is instructive to examine two mechanisms that could ensure the invariance of preferences: canonical representations and the use of expected actuarial value.

Invariance would hold if all formulations of the same prospect were transformed to a standard canonical representation (e.g., a cumulative probability distribution of the same random variable) because the various versions would then all be evaluated in the same manner. In problem 2, for example, invariance and dominance would both be preserved if the outcomes of the two decisions were aggregated prior to evaluation. Similarly, the same choice would be made in both versions of the medical problem if the outcomes were coded in terms of one dominant frame (e.g., rate of survival). The observed failures of invariance indicate that people do not spontaneously aggregate concurrent prospects or transform all outcomes into a common frame.

The failure to construct a canonical representation in decision problems contrasts with other cognitive tasks in which such representations are generated automatically and effortlessly. In particular, our visual experience consists largely of canonical representations: objects do not appear to change in size, shape, brightness, or color when we move around them or when illumination varies. A white circle seen from a sharp angle in dim light appears circular and white, not ellipsoid and grey. Canonical representations are also generated in the process of language comprehension, where listeners quickly recode much of what they hear into an abstract propositional form that no longer discriminates, for example, between the active and the passive voice and often does not distinguish what was actually said from what was implied or presupposed (Clark and Clark 1977). Unfortunately, the mental machinery that transforms percepts and sentences into standard forms does not automatically apply to the process of choice.

Invariance could be satisfied even in the absence of a canonical representation if the evaluation of prospects were separately linear, or nearly linear, in probability and monetary value. If people ordered risky prospects by their actuarial values, invariance and dominance would always hold. In particular, there would be no difference between the mortality and the survival versions of the medical problem. Because the evaluation of outcomes and probabilities is generally nonlinear, and because people do not spontaneously construct canonical representations of decisions, invariance commonly fails. Normative models of choice, which assume invariance, therefore cannot provide an adequate descriptive account of choice behavior. In the next section we present a descriptive account of risky choice, called prospect theory, and explore its consequences. Failures of invariance are explained by framing effects that control the representation of options, in conjunction with the nonlinearities of value and belief.

III. Framing and Evaluation of Outcomes

Prospect theory distinguishes two phases in the choice process: a phase of framing and editing, followed by a phase of evaluation (Kahneman and Tversky 1979). The first phase consists of a preliminary analysis of the decision problem, which frames the effective acts, contingencies, and outcomes. Framing is controlled by the manner in which the choice problem is presented as well as by norms, habits, and expectancies of the decision maker. Additional operations that are performed prior to evaluation include cancellation of common components and the elimination of options that are seen to be dominated by others. In the second phase, the framed prospects are evaluated, and the prospect of highest value is selected. The theory distinguishes two ways of choosing between prospects: by detecting that one dominates another or by comparing their values.

For simplicity, we confine the discussion to simple gambles with numerical probabilities and monetary outcomes. Let $(x, p; y, q)$ denote a prospect that yields x with probability p and y with probability q and that preserves the status quo with probability $(1 - p - q)$. According to prospect theory, there are values $v(\cdot)$, defined on gains and losses, and decision weights $\pi(\cdot)$, defined on stated probabilities, such that the overall value of the prospect equals $\pi(p)v(x) + \pi(q)v(y)$. A slight modification is required if all outcomes of a prospect have the same sign.[2]

2. If $p + q = 1$ and either $x > y > 0$ or $x < y < 0$, the value of a prospect is given by $v(y) + \pi(p)[v(x) - v(y)]$, so that decision weights are not applied to sure outcomes.

The Value Function

Following Markowitz (1952), outcomes are expressed in prospect theory as positive or negative deviations (gains or losses) from a neutral reference outcome, which is assigned a value of zero. Unlike Markowitz, however, we propose that the value function is commonly S shaped, concave above the reference point, and convex below it, as illustrated in figure 1. Thus the difference in subjective value between a gain of $100 and a gain of $200 is greater than the subjective difference between a gain of $1,100 and a gain of $1,200. The same relation between value differences holds for the corresponding losses. The proposed function expresses the property that the effect of a marginal change decreases with the distance from the reference point in either direction. These hypotheses regarding the typical shape of the value function may not apply to ruinous losses or to circumstances in which particular amounts assume special significance.

A significant property of the value function, called *loss aversion*, is that the response to losses is more extreme than the response to gains. The common reluctance to accept a fair bet on the toss of a coin suggests that the displeasure of losing a sum of money exceeds the pleasure of winning the same amount. Thus the proposed value function is (i) defined on gains and losses, (ii) generally concave for gains and convex for losses, and (iii) steeper for losses than for gains. These properties of the value function have been supported in many studies of risky choice involving monetary outcomes (Fishburn and Kochenberger 1979; Kahneman and Tversky 1979; Hershey and Schoemaker 1980; Payne, Laughhunn, and Crum 1980) and human lives (Tversky 1977; Eraker and Sox 1981; Tversky and Kahneman 1981; Fischhoff 1983). Loss aversion may also contribute to the observed discrepancies between the amount of money people are willing to pay for a good and the compensation they demand to give it up (Bishop and Heberlein 1979; Knetsch and Sinden 1984). This effect is implied by the value function if the good is valued as a gain in the former context and as a loss in the latter.

Framing Outcomes

The framing of outcomes and the contrast between traditional theory and the present analysis are illustrated in the following problems.

Problem 3 (N = 126): Assume yourself richer by $300 than you are today. You have to choose between
a sure gain of $100 [72%]
50% chance to gain $200 and 50% chance to gain nothing [28%]

Problem 4 (N = 128): Assume yourself richer by $500 than you are today. You have to choose between
a sure loss of $100 [36%]
50% chance to lose nothing and 50% chance to lose $200 [64%]

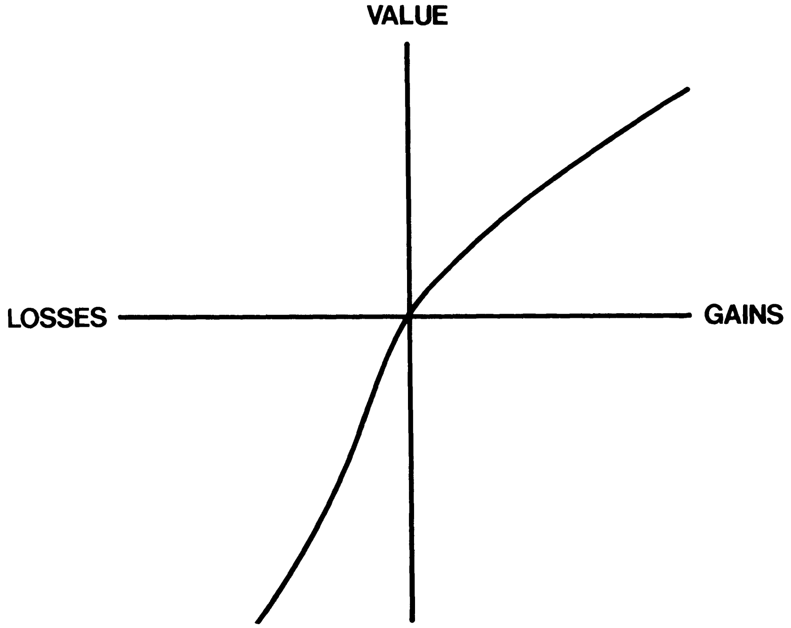


FIG. 1.—A typical value function

As implied by the value function, the majority choice is risk averse in problem 3 and risk seeking in problem 4, although the two problems are essentially identical. In both cases one faces a choice between $400 for sure and an even chance of $500 or $300. Problem 4 is obtained from problem 3 by increasing the initial endowment by $200 and subtracting this amount from both options. This variation has a substantial effect on preferences. Additional questions showed that variations of $200 in initial wealth have little or no effect on choices. Evidently, preferences are quite insensitive to small changes of wealth but highly sensitive to corresponding changes in reference point. These observations show that the effective carriers of values are gains and losses, or changes in wealth, rather than states of wealth as implied by the rational model.

The common pattern of preferences observed in problems 3 and 4 is of special interest because it violates not only expected utility theory but practically all other normatively based models of choice. In particular, these data are inconsistent with the model of regret advanced by Bell (1982) and by Loomes and Sugden (1982) and axiomatized by Fishburn (1982). This follows from the fact that problems 3 and 4 yield identical outcomes and an identical regret structure. Furthermore, regret theory cannot accommodate the combination of risk aversion in problem 3 and risk seeking in problem 4—even without the corresponding changes in endowment that make the problems extensionally equivalent.

Shifts of reference can be induced by different decompositions of outcomes into risky and riskless components, as in the above problems. The reference point can also be shifted by a mere labeling of outcomes, as illustrated in the following problems (Tversky and Kahneman 1981, p. 453).

> Problem 5 ($N = 152$): Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:
>
> If Program A is adopted, 200 people will be saved. [72%]
>
> If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [28%]

In problem 5 the outcomes are stated in positive terms (lives saved), and the majority choice is accordingly risk averse. The prospect of certainly saving 200 lives is more attractive than a risky prospect of equal expected value. A second group of respondents was given the same cover story with the following descriptions of the alternative programs.

> Problem 6 ($N = 155$):
> If Program C is adopted 400 people will die. [22%]
>
> If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. [78%]

In problem 6 the outcomes are stated in negative terms (lives lost), and the majority choice is accordingly risk seeking. The certain death of 400 people is less acceptable than a two-thirds chance that 600 people will die. Problems 5 and 6, however, are essentially identical. They differ only in that the former is framed in terms of the number of lives saved (relative to an expected loss of 600 lives if no action is taken), whereas the latter is framed in terms of the number of lives lost.

On several occasions we presented both versions to the same respondents and discussed with them the inconsistent preferences evoked by the two frames. Many respondents expressed a wish to remain risk averse in the "lives saved" version and risk seeking in the "lives lost" version, although they also expressed a wish for their answers to be consistent. In the persistence of their appeal, framing effects resemble visual illusions more than computational errors.

Discounts and Surcharges

Perhaps the most distinctive intellectual contribution of economic analysis is the systematic consideration of alternative opportunities. A basic principle of economic thinking is that opportunity costs and out-of-pocket costs should be treated alike. Preferences should depend only on relevant differences between options, not on how these differences are labeled. This principle runs counter to the psychological tendencies that make preferences susceptible to superficial variations in form. In particular, a difference that favors outcome A over outcome B can sometimes be framed either as an advantage of A or as a disadvantage of B by suggesting either B or A as the neutral reference point. Because of loss aversion, the difference will loom larger when A is neutral and B-A is evaluated as a loss than when B is neutral and A-B is evaluated as a gain. The significance of such variations of framing has been noted in several contexts.

Thaler (1980) drew attention to the effect of labeling a difference between two prices as a surcharge or a discount. It is easier to forgo a discount than to accept a surcharge because the same price difference is valued as a gain in the former case and as a loss in the latter. Indeed, the credit card lobby is said to insist that any price difference between cash and card purchases should be labeled a cash discount rather than a credit surcharge. A similar idea could be invoked to explain why the price response to slack demand often takes the form of discounts or special concessions (Stigler and Kindahl 1970). Customers may be expected to show less resistance to the eventual cancellation of such temporary arrangements than to outright price increases. Judgments of fairness exhibit the same pattern (Kahneman, Knetsch, and Thaler, in this issue).

Schelling (1981) has described a striking framing effect in a context of tax policy. He points out that the tax table can be constructed by using as a default case either the childless family (as is in fact done) or, say, the modal two-child family. The tax difference between a childless family and a two-child family is naturally framed as an exemption (for the two-child family) in the first frame and as a tax premium (on the childless family) in the second frame. This seemingly innocuous difference has a large effect on judgments of the desired relation between income, family size, and tax. Schelling reported that his students rejected the idea of granting the rich a larger exemption than the poor in the first frame but favored a larger tax premium on the childless rich than on the childless poor in the second frame. Because the exemption and the premium are alternative labels for the same tax differences in the two cases, the judgments violate invariance. Framing the consequences of a public policy in positive or in negative terms can greatly alter its appeal.

The notion of a money illusion is sometimes applied to workers' willingness to accept, in periods of high inflation, increases in nominal wages that do not protect their real income—although they would strenuously resist equivalent wage cuts in the absence of inflation. The essence of the illusion is that, whereas a cut in the nominal wage is always recognized as a loss, a nominal increase that does not preserve real income may be treated as a gain. Another manifestation of the money illusion was observed in a study of the perceived fairness of economic actions (Kahneman, Knetsch, and Thaler, in press). Respondents in a telephone interview evaluated the fairness of the action described in the following vignette, which was presented in two versions that differed only in the bracketed clauses.

> A company is making a small profit. It is located in a community experiencing a recession with substantial unemployment [but no inflation/and inflation of 12%]. The company decides to [decrease wages and salaries 7%/increase salaries only 5%] this year.

Although the loss of real income is very similar in the two versions, the proportion of respondents who judged the action of the company "unfair" or "very unfair" was 62% for a nominal reduction but only 22% for a nominal increase.

Bazerman (1983) has documented framing effects in experimental studies of bargaining. He compared the performance of experimental subjects when the outcomes of bargaining were formulated as gains or as losses. Subjects who bargained over the allocation of losses more often failed to reach agreement and more often failed to discover a Pareto-optimal solution. Bazerman attributed these observations to the general propensity toward risk seeking in the domain of losses, which may increase the willingness of both participants to risk the negative consequences of a deadlock.

Loss aversion presents an obstacle to bargaining whenever the participants evaluate their own concessions as losses and the concessions obtained from the other party as gains. In negotiating over missiles, for example, the subjective loss of security associated with dismantling a missile may loom larger than the increment of security produced by a similar action on the adversary's part. If the two parties both assign a two-to-one ratio to the values of the concessions they make and of those they obtain, the resulting four-to-one gap may be difficult to bridge. Agreement will be much easier to achieve by negotiators who trade in "bargaining chips" that are valued equally, regardless of whose hand they are in. In this mode of trading, which may be common in routine purchases, loss aversion tends to disappear (Kahneman and Tversky 1984).

IV. The Framing and Weighting of Chance Events

In expected-utility theory, the utility of each possible outcome is weighted by its probability. In prospect theory, the value of an uncertain outcome is multiplied by a decision weight $\pi(p)$, which is a monotonic function of p but is not a probability. The weighting function π

has the following properties. First, impossible events are discarded, that is, $\pi(0) = 0$, and the scale is normalized so that $\pi(1) = 1$, but the function is not well behaved near the end points (Kahneman and Tversky 1979). Second, for low probabilities, $\pi(p) > p$, but $\pi(p) + \pi(1 - p) \leq 1$ (subcertainty). Thus low probabilities are overweighted, moderate and high probabilities are underweighted, and the latter effect is more pronounced than the former. Third, $\pi(pr)/\pi(p) < \pi(pqr)/\pi(pq)$ for all $0 < p, q, r \leq 1$ (subproportionality). That is, for any fixed probability ratio r, the ratio of decision weights is closer to unity when the probabilities are low than when they are high, for example, $\pi(.1)/\pi(.2) > \pi(.4)/\pi(.8)$. A hypothetical weighting function that satisfies these properties is shown in figure 2. Its consequences are discussed in the next section.[3]

Nontransparent Dominance

The major characteristic of the weighting function is the overweighting of probability differences involving certainty and impossibility, for example, $\pi(1.0) - \pi(.9)$ or $\pi(.1) - \pi(0)$, relative to comparable differences in the middle of the scale, for example, $\pi(.3) - \pi(.2)$. In particular, for small p, π is generally subadditive, for example, $\pi(.01) + \pi(.06) > \pi(.07)$. This property can lead to violations of dominance, as illustrated in the following pair of problems.

> Problem 7 ($N = 88$). Consider the following two lotteries, described by the percentage of marbles of different colors in each box and the amount of money you win or lose depending on the color of a randomly drawn marble. Which lottery do you prefer?

Option A					
	90% white	6% red	1% green	1% blue	2% yellow
	$0	win $45	win $30	lose $15	lose $15
Option B					
	90% white	6% red	1% green	1% blue	2% yellow
	$0	win $45	win $45	lose $10	lose $15

It is easy to see that option B dominates option A: for every color the outcome of B is at least as desirable as the outcome of A. Indeed, all

3. The extension of the present analysis to prospects with many (nonzero) outcomes involves two additional steps. First, we assume that continuous (or multivalued) distributions are approximated, in the framing phase, by discrete distributions with a relatively small number of outcomes. For example, a uniform distribution on the interval (0, 90) may be represented by the discrete prospect (0, .1; 10, .1; . . . ; 90, .1). Second, in the multiple-outcome case the weighting function, $\pi_p(p_i)$, must depend on the probability vector p, not only on the component p_i, $i = 1, \ldots, n$. For example, Quiggin (1982) uses the function $\pi_p(p_i) = \pi(p_i)/[\pi(p_1) + \ldots + \pi(p_n)]$. As in the two-outcome case, the weighting function is assumed to satisfy subcertainty, $\pi_p(p_1) + \ldots + \pi_p(p_n) \leq 1$, and subproportionality.

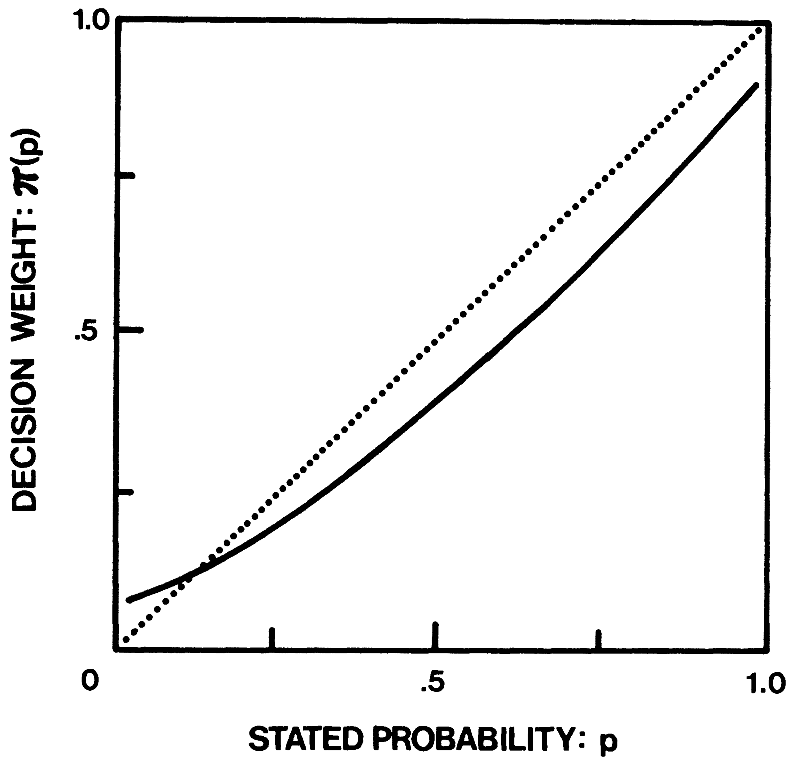


FIG. 2.—A typical weighting function

respondents chose B over A. This observation is hardly surprising because the relation of dominance is highly transparent, so the dominated prospect is rejected without further processing. The next problem is effectively identical to problem 7, except that colors yielding identical outcomes (red and green in B and yellow and blue in A) are combined. We have proposed that this operation is commonly performed by the decision maker if no dominated prospect is detected.

Problem 8 (N = 124). Which lottery do you prefer?

Option C

90% white	6% red	1% green	3% yellow
$0	win $45	win $30	lose $15

Option D

90% white	7% red	1% green	2% yellow
$0	win $45	lose $10	lose $15

The formulation of problem 8 simplifies the options but masks the relation of dominance. Furthermore, it enhances the attractiveness of

C, which has two positive outcomes and one negative, relative to D, which has two negative outcomes and one positive. As an inducement to consider the options carefully, participants were informed that one-tenth of them, selected at random, would actually play the gambles they chose. Although this announcement aroused much excitement, 58% of the participants chose the dominated alternative C. In answer to another question the majority of respondents also assigned a higher cash equivalent to C than to D. These results support the following propositions. (i) Two formulations of the same problem elicit different preferences, in violation of invariance. (ii) The dominance rule is obeyed when its application is transparent. (iii) Dominance is masked by a frame in which the inferior option yields a more favorable outcome in an identified state of the world (e.g., drawing a green marble). (iv) The discrepant preferences are consistent with the subadditivity of decision weights. The role of transparency may be illuminated by a perceptual example. Figure 3 presents the well-known Müller-Lyer illusion: the top line appears longer than the bottom line, although it is in fact shorter. In figure 4, the same patterns are embedded in a rectangular frame, which makes it apparent that the protruding bottom line is longer than the top one. This judgment has the nature of an inference, in contrast to the perceptual impression that mediates judgment in figure 3. Similarly, the finer partition introduced in problem 7 makes it possible to conclude that option D is superior to C, without assessing their values. Whether the relation of dominance is detected depends on framing as well as on the sophistication and experience of the decision maker. The dominance relation in problems 8 and 1 could be transparent to a sophisticated decision maker, although it was not transparent to most of our respondents.

Certainty and Pseudocertainty

The overweighting of outcomes that are obtained with certainty relative to outcomes that are merely probable gives rise to violations of the expectation rule, as first noted by Allais (1953). The next series of problems (Tversky and Kahneman 1981, p. 455) illustrates the phenomenon discovered by Allais and its relation to the weighting of probabilities and to the framing of chance events. Chance events were realized by drawing a single marble from a bag containing a specified number of favorable and unfavorable marbles. To encourage thoughtful answers, one-tenth of the participants, selected at random, were given an opportunity to play the gambles they chose. The same respondents answered problems 9–11, in that order.

Problem 9 ($N = 77$). Which of the following options do you prefer?
A. a sure gain of $30 [78%]
B. 80% chance to win $45 and 20% chance to win nothing [22%]

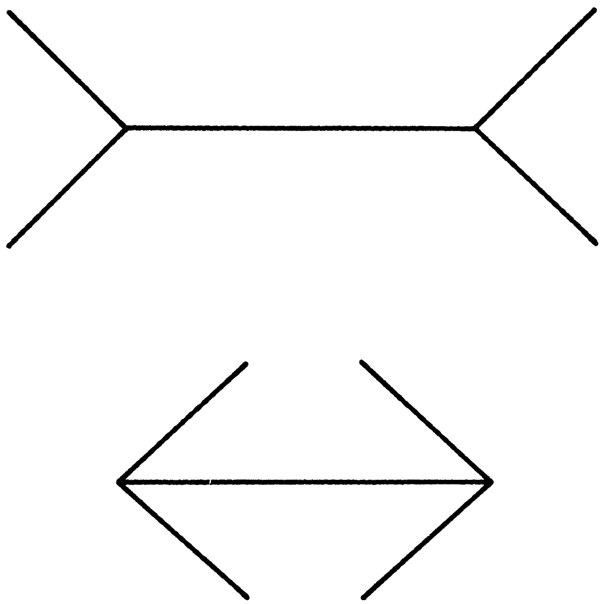


FIG. 3.—The Müller-Lyer illusion

Problem 10 ($N = 81$). Which of the following options do you prefer?
C. 25% chance to win $30 and 75% chance to win nothing [42%]
D. 20% chance to win $45 and 80% chance to win nothing [58%]

Note that problem 10 is obtained from problem 9 by reducing the probabilities of winning by a factor of four. In expected utility theory a preference for A over B in problem 9 implies a preference for C over D in problem 10. Contrary to this prediction, the majority preference switched from the lower prize ($30) to the higher one ($45) when the probabilities of winning were substantially reduced. We called this phenomenon the *certainty effect* because the reduction of the probability of winning from certainty to .25 has a greater effect than the corresponding reduction from .8 to .2. In prospect theory, the modal choice in problem 9 implies $v(45)\pi(.80) < v(30)\pi(1.0)$, whereas the modal choice in problem 10 implies $v(45)\pi(.20) > v(30)\pi(.25)$. The observed violation of expected utility theory, then, is implied by the curvature of π (see fig. 2) if

$$\frac{\pi(.20)}{\pi(.25)} > \frac{v(30)}{v(45)} > \frac{\pi(.80)}{\pi(1.0)}.$$

Allais's problem has attracted the attention of numerous theorists, who attempted to provide a normative rationale for the certainty effect by relaxing the cancellation rule (see, e.g., Allais 1979; Fishburn 1982, 1983; Machina 1982; Quiggin 1982; Chew 1983). The following problem

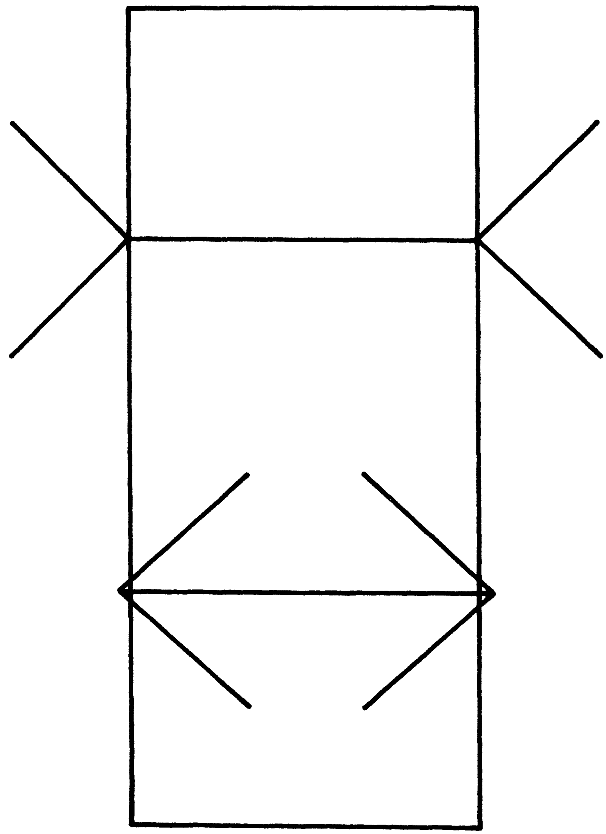


FIG. 4.—A transparent version of the Müller-Lyer illusion

illustrates a related phenomenon, called the *pseudocertainty effect*, that cannot be accommodated by relaxing cancellation because it also involves a violation of invariance.

Problem 11 ($N = 85$): Consider the following two stage game. In the first stage, there is a 75% chance to end the game without winning anything, and a 25% chance to move into the second stage. If you reach the second stage you have a choice between:

E. a sure win of $30 [74%]
F. 80% chance to win $45 and 20% chance to win nothing [26%]

Your choice must be made before the outcome of the first stage is known.

Because there is one chance in four to move into the second stage, prospect E offers a .25 probability of winning $30, and prospect F offers a .25 × .80 = .20 probability of winning $45. Problem 11 is therefore identical to problem 10 in terms of probabilities and outcomes. However, the preferences in the two problems differ: most subjects made a risk-averse choice in problem 11 but not in problem 10. We call this phenomenon the pseudocertainty effect because an outcome that is actually uncertain is weighted as if it were certain. The framing of problem 11 as a two-stage game encourages respondents to apply cancellation: the event of failing to reach the second stage is discarded prior to evaluation because it yields the same outcomes in both options. In this framing problems 11 and 9 are evaluated alike.

Although problems 10 and 11 are identical in terms of final outcomes and their probabilities, problem 11 has a greater potential for inducing regret. Consider a decision maker who chooses F in problem 11, reaches the second stage, but fails to win the prize. This individual knows that the choice of E would have yielded a gain of $30. In problem 10, on the other hand, an individual who chooses D and fails to win cannot know with certainty what the outcome of the other choice would have been. This difference could suggest an alternative interpretation of the pseudocertainty effect in terms of regret (e.g., Loomes and Sugden 1982). However, the certainty and the pseudocertainty effects were found to be equally strong in a modified version of problems 9–11 in which opportunities for regret were equated across problems. This finding does not imply that considerations of regret play no role in decisions. (For examples, see Kahneman and Tversky [1982, p. 710].) It merely indicates that Allais's example and the pseudocertainty effect are primarily controlled by the nonlinearity of decision weights and the framing of contingencies rather than by the anticipation of regret.[4]

The certainty and pseudocertainty effects are not restricted to monetary outcomes. The following problem illustrates these phenomena in a medical context. The respondents were 72 physicians attending a meeting of the California Medical Association. Essentially the same pattern of responses was obtained from a larger group ($N = 180$) of college students.

> Problem 12 ($N = 72$). In the treatment of tumors there is sometimes a choice between two types of therapies: (i) a radical treatment such as extensive surgery, which involves some risk of imminent death,

4. In the modified version—problems 9′–11′—the probabilities of winning were generated by drawing a number from a bag containing 100 sequentially numbered tickets. In problem 10′, the event associated with winning $45 (drawing a number between one and 20) was included in the event associated with winning $30 (drawing a number between one and 25). The sequential setup of problem 11 was replaced by the simultaneous play of two chance devices: the roll of a die (whose outcome determines whether the game is on) and the drawing of a numbered ticket from a bag. The possibility of regret now exists in all three problems, and problem 10′ and 11′ no longer differ in this respect because a decision maker would always know the outcomes of alternative choices. Consequently, regret theory cannot explain either the certainty effect (9′ vs. 10′) or the pseudocertainty effect (10′ vs. 11′) observed in the modified problems.

(ii) a moderate treatment, such as limited surgery or radiation therapy. Each of the following problems describes the possible outcome of two alternative treatments, for three different cases. In considering each case, suppose the patient is a 40-year-old male. Assume that without treatment death is imminent (within a month) and that only one of the treatments can be applied. Please indicate the treatment you would prefer in each case.

Case 1

Treatment A: 20% chance of imminent death and 80% chance of normal life, with an expected longevity of 30 years. [35%]

Treatment B: certainty of a normal life, with an expected longevity of 18 years. [65%]

Case 2

Treatment C: 80% chance of imminent death and 20% chance of normal life, with an expected longevity of 30 years. [68%]

Treatment D: 75% chance of imminent death and 25% chance of normal life, with an expected longevity of 18 years. [32%]

Case 3

Consider a new case where there is a 25% chance that the tumor is treatable and a 75% chance that it is not. If the tumor is not treatable, death is imminent. If the tumor is treatable, the outcomes of the treatment are as follows:

Treatment E: 20% chance of imminent death and 80% chance of normal life, with an expected longevity of 30 years. [32%]

Treatment F: certainty of normal life, with an expected longevity of 18 years. [68%]

The three cases of this problem correspond, respectively, to problems 9–11, and the same pattern of preferences is observed. In case 1, most respondents make a risk-averse choice in favor of certain survival with reduced longevity. In case 2, the moderate treatment no longer ensures survival, and most respondents choose the treatment that offers the higher expected longevity. In particular, 64% of the physicians who chose B in case 1 selected C in case 2. This is another example of Allais's certainty effect.

The comparison of cases 2 and 3 provides another illustration of pseudocertainty. The cases are identical in terms of the relevant outcomes and their probabilities, but the preferences differ. In particular, 56% of the physicians who chose C in case 2 selected F in case 3. The conditional framing induces people to disregard the event of the tumor not being treatable because the two treatments are equally ineffective in this case. In this frame, treatment F enjoys the advantage of pseudocertainty. It appears to ensure survival, but the assurance is conditional on the treatability of the tumor. In fact, there is only a .25 chance of surviving a month if this option is chosen.

The conjunction of certainty and pseudocertainty effects has significant implications for the relation between normative and descriptive theories of choice. Our results indicate that cancellation is actually obeyed in choices—in those problems that make its application transparent. Specifically, we find that people make the same choices in problems 11 and 9 and in cases 3 and 1 of problem 12. Evidently, people "cancel" an event that yields the same outcomes for all options, in two-stage or nested structures. Note that in these examples cancellation is satisfied in problems that are formally equivalent to those in which it is violated. The empirical validity of cancellation therefore depends on the framing of the problems.

The present concept of framing originated from the analysis of Allais's problems by Savage (1954, pp. 101–4) and Raiffa (1968, pp. 80–86), who reframed these examples in an attempt to make the application of cancellation more compelling. Savage and Raiffa were right: naive respondents indeed obey the cancellation axiom when its application is sufficiently transparent.[5] However, the contrasting preferences in different versions of the same choice (problems 10 and 11 and cases 2 and 3 of problem 12) indicate that people do not follow the same axiom when its application is not transparent. Instead, they apply (nonlinear) decision weights to the probabilities as stated. The status of cancellation is therefore similar to that of dominance: both rules are intuitively compelling as abstract principles of choice, consistently obeyed in transparent problems and frequently violated in nontransparent ones. Attempts to rationalize the preferences in Allais's example by discarding the cancellation axiom face a major difficulty: they do not distinguish transparent formulations in which cancellation is obeyed from nontransparent ones in which it is violated.

V. Discussion

In the preceding sections we challenged the descriptive validity of the major tenets of expected utility theory and outlined an alternative account of risky choice. In this section we discuss alternative theories

5. It is noteworthy that the conditional framing used in problems 11 and 12 (case 3) is much more effective in eliminating the common responses to Allais's paradox than the partition framing introduced by Savage (see, e.g., Slovic and Tversky 1974). This is probably due to the fact that the conditional framing makes it clear that the critical options are identical—after eliminating the state whose outcome does not depend on one's choice (i.e., reaching the second stage in problem 11, an untreatable tumor in problem 12, case 3).

and argue against the reconciliation of normative and descriptive analyses. Some objections of economists to our analysis and conclusions are addressed.

Descriptive and Normative Considerations

Many alternative models of risky choice, designed to explain the observed violations of expected utility theory, have been developed in the last decade. These models divide into the following four classes. (i) Nonlinear functionals (e.g., Allais 1953, 1979; Machina 1982) are obtained by eliminating the cancellation condition altogether. These models do not have axiomatizations leading to a (cardinal) measurement of utility, but they impose various restrictions (i.e., differentiability) on the utility functional. (ii) The expectations quotient model (axiomatized by Chew and MacCrimmon 1979; Weber 1982; Chew 1983; Fishburn 1983) replaces cancellation by a weaker substitution axiom and represents the value of a prospect by the ratio of two linear functionals. (iii) Bilinear models with nonadditive probabilities (e.g., Kahneman and Tversky 1979; Quiggin 1982; Schmeidler 1984; Segal 1984; Yaari 1984; Luce and Narens 1985) assume various restricted versions of cancellation (or substitution) and construct a bilinear representation in which the utilities of outcomes are weighted by a nonadditive probability measure or by some nonlinear transform of the probability scale. (iv) Nontransitive models represent preferences by a bivariate utility function. Fishburn (1982, 1984) axiomatized such models, while Bell (1982) and Loomes and Sugden (1982) interpreted them in terms of expected regret. For further theoretical developments, see Fishburn (1985).

The relation between models and data is summarized in table 1. The stub column lists the four major tenets of expected utility theory. Column 1 lists the major empirical violations of these tenets and cites a few representative references. Column 2 lists the subset of models discussed above that are consistent with the observed violations.

TABLE 1 **Summary of Empirical Violations and Explanatory Models**

Tenet	Empirical Violation	Explanatory Model
Cancellation	Certainty effect (Allais 1953, 1979; Kahneman and Tversky 1979) (problems 9–10, and 12 [cases 1 and 2])	All models
Transitivity	Lexicographic semiorder (Tversky 1969) Preference reversals (Slovic and Lichtenstein 1983)	Bivariate models
Dominance	Contrasting risk attitudes (problem 2) Subadditive decision weights (problem 8)	Prospect theory
Invariance	Framing effects (Problems 1, 3–4, 5–6, 7–8, 10–11, and 12)	Prospect theory

The conclusions of table 1 may be summarized as follows. First, all the above models (as well as some others) are consistent with the violations of cancellation produced by the certainty effect.[6] Therefore, Allais's "paradox" cannot be used to compare or evaluate competing nonexpectation models. Second, bivariate (nontransitive) models are needed to explain observed intransitivities. Third, only prospect theory can accommodate the observed violations of (stochastic) dominance and invariance. Although some models (e.g., Loomes and Sugden 1982; Luce and Narens 1985) permit some limited failures of invariance, they do not account for the range of framing effects described in this article.

Because framing effects and the associated failures of invariance are ubiquitous, no adequate descriptive theory can ignore these phenomena. On the other hand, because invariance (or extensionality) is normatively indispensable, no adequate prescriptive theory should permit its violation. Consequently, the dream of constructing a theory that is acceptable both descriptively and normatively appears unrealizable (see also Tversky and Kahneman 1983).

Prospect theory differs from the other models mentioned above in being unabashedly descriptive and in making no normative claims. It is designed to explain preferences, whether or not they can be rationalized. Machina (1982, p. 292) claimed that prospect theory is "unacceptable as a descriptive model of behavior toward risk" because it implies violations of stochastic dominance. But since the violations of dominance predicted by the theory have actually been observed (see problems 2 and 8), Machina's objection appears invalid.

Perhaps the major finding of the present article is that the axioms of rational choice are generally satisfied in transparent situations and often violated in nontransparent ones. For example, when the relation of stochastic dominance is transparent (as in the aggregated version of problem 2 and in problem 7), practically everyone selects the dominant prospect. However, when these problems are framed so that the relation of dominance is no longer transparent (as in the segregated version of problem 2 and in problem 8), most respondents violate dominance, as predicted. These results contradict all theories that imply stochastic dominance as well as others (e.g., Machina 1982) that predict the same choices in transparent and nontransparent contexts. The same conclusion applies to cancellation, as shown in the discussion of pseudocertainty. It appears that both cancellation and dominance have normative appeal, although neither one is descriptively valid.

The present results and analysis—particularly the role of transparency and the significance of framing—are consistent with the concep-

6. Because the present article focuses on prospects with known probabilities, we do not discuss the important violations of cancellation due to ambiguity (Ellsberg 1961).

tion of bounded rationality originally presented by Herbert Simon (see, e.g., Simon 1955, 1978; March 1978; Nelson and Winter 1982). Indeed, prospect theory is an attempt to articulate some of the principles of perception and judgment that limit the rationality of choice.

The introduction of psychological considerations (e.g., framing) both enriches and complicates the analysis of choice. Because the framing of decisions depends on the language of presentation, on the context of choice, and on the nature of the display, our treatment of the process is necessarily informal and incomplete. We have identified several common rules of framing, and we have demonstrated their effects on choice, but we have not provided a formal theory of framing. Furthermore, the present analysis does not account for all the observed failures of transitivity and invariance. Although some intransitivities (e.g., Tversky 1969) can be explained by discarding small differences in the framing phase, and others (e.g., Raiffa 1968, p. 75) arise from the combination of transparent and nontransparent comparisons, there are examples of cyclic preferences and context effects (see, e.g., Slovic, Fischhoff, and Lichtenstein 1982; Slovic and Lichtenstein 1983) that require additional explanatory mechanisms (e.g., multiple reference points and variable weights). An adequate account of choice cannot ignore these effects of framing and context, even if they are normatively distasteful and mathematically intractable.

Bolstering Assumptions

The assumption of rationality has a favored position in economics. It is accorded all the methodological privileges of a self-evident truth, a reasonable idealization, a tautology, and a null hypothesis. Each of these interpretations either puts the hypothesis of rational action beyond question or places the burden of proof squarely on any alternative analysis of belief and choice. The advantage of the rational model is compounded because no other theory of judgment and decision can ever match it in scope, power, and simplicity.

Furthermore, the assumption of rationality is protected by a formidable set of defenses in the form of bolstering assumptions that restrict the significance of any observed violation of the model. In particular, it is commonly assumed that substantial violations of the standard model are (i) restricted to insignificant choice problems, (ii) quickly eliminated by learning, or (iii) irrelevant to economics because of the corrective function of market forces. Indeed, incentives sometimes improve the quality of decisions, experienced decision makers often do better than novices, and the forces of arbitrage and competition can nullify some effects of error and illusion. Whether these factors ensure rational choices in any particular situation is an empirical issue, to be settled by observation, not by supposition.

It has frequently been claimed (see, e.g., Smith 1985) that the observed failures of rational models are attributable to the cost of thinking and will thus be eliminated by proper incentives. Experimental findings provide little support for this view. Studies reported in the economic and psychological literature have shown that errors that are prevalent in responses to hypothetical questions persist even in the presence of significant monetary payoffs. In particular, elementary blunders of probabilistic reasoning (Grether 1980; Tversky and Kahneman 1983), major inconsistencies of choice (Grether and Plott 1979; Slovic and Lichtenstein 1983), and violations of stochastic dominance in nontransparent problems (see problem 2 above) are hardly reduced by incentives. The evidence that high stakes do not always improve decisions is not restricted to laboratory studies. Significant errors of judgment and choice can be documented in real world decisions that involve high stakes and serious deliberation. The high rate of failures of small businesses, for example, is not easily reconciled with the assumptions of rational expectations and risk aversion.

Incentives do not operate by magic: they work by focusing attention and by prolonging deliberation. Consequently, they are more likely to prevent errors that arise from insufficient attention and effort than errors that arise from misperception or faulty intuition. The example of visual illusion is instructive. There is no obvious mechanism by which the mere introduction of incentives (without the added opportunity to make measurements) would reduce the illusion observed in figure 3, and the illusion vanishes—even in the absence of incentives—when the display is altered in figure 4. The corrective power of incentives depends on the nature of the particular error and cannot be taken for granted.

The assumption of the rationality of decision making is often defended by the argument that people will learn to make correct decisions and sometimes by the evolutionary argument that irrational decision makers will be driven out by rational ones. There is no doubt that learning and selection do take place and tend to improve efficiency. As in the case of incentives, however, no magic is involved. Effective learning takes place only under certain conditions: it requires accurate and immediate feedback about the relation between the situational conditions and the appropriate response. The necessary feedback is often lacking for the decisions made by managers, entrepreneurs, and politicians because (i) outcomes are commonly delayed and not easily attributable to a particular action; (ii) variability in the environment degrades the reliability of the feedback, especially where outcomes of low probability are involved; (iii) there is often no information about what the outcome would have been if another decision had been taken; and (iv) most important decisions are unique and therefore provide little opportunity for learning (see Einhorn and Hogarth 1978). The conditions for organizational learning are hardly better. Learning surely occurs, for both individuals and organizations, but any claim that a particular error will be eliminated by experience must be supported by demonstrating that the conditions for effective learning are satisfied.

Finally, it is sometimes argued that failures of rationality in individual decision making are inconsequential because of the corrective effects of the market (Knez, Smith, and Williams 1985). Economic agents are often protected from their own irrational predilections by the forces of competition and by the action of arbitrageurs, but there are situations in which this mechanism fails. Hausch, Ziemba, and Rubenstein (1981) have documented an instructive example: the market for win bets at the racetrack is efficient, but the market for bets on place and show is not. Bettors commonly underestimate the probability that the favorite will end up in second or third place, and this effect is sufficiently large to sustain a contrarian betting strategy with a positive expected value. This inefficiency is found in spite of the high incentives, of the unquestioned level of dedication and expertise among participants in racetrack markets, and of obvious opportunities for learning and for arbitrage.

Situations in which errors that are common to many individuals are unlikely to be corrected by the market have been analyzed by Haltiwanger and Waldman (1985) and by Russell and Thaler (1985). Furthermore, Akerlof and Yellen (1985) have presented their near-rationality theory, in which some prevalent errors in responding to economic changes (e.g., inertia or money illusion) will (i) have little effect on the individual (thereby eliminating the possibility of learning), (ii) provide no opportunity for arbitrage, and yet (iii) have large economic effects. The claim that the market can be trusted to correct the effect of individual irrationalities cannot be made without supporting evidence, and the burden of specifying a plausible corrective mechanism should rest on those who make this claim.

The main theme of this article has been that the normative and the descriptive analyses of choice should be viewed as separate enterprises. This conclusion suggests a research agenda. To retain the rational model in its customary descriptive role, the relevant bolstering assumptions must be validated. Where these assumptions fail, it is instructive to trace the implications of the descriptive analysis (e.g., the effects of loss aversion, pseudocertainty, or the money illusion) for public policy, strategic decision making, and macroeconomic phenomena (see Arrow 1982; Akerlof and Yellen 1985).

References

Akerlof, G. A., and Yellen, J. 1985. Can small deviations from rationality make significant differences to economic equilibria? *American Economic Review* 75:708–20.

Allais, M. 1953. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'Ecole Américaine. *Econometrica* 21:503–46.

Allais, M. 1979. The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American School. In M. Allais and O. Hagen (eds.), *Expected Utility Hypotheses and the Allais Paradox*. Dordrecht: Reidel.
Arrow, K. J. 1982. Risk perception in psychology and economics. *Economic Inquiry* 20:1–9.
Bazerman, M. H. 1983. Negotiator judgment. *American Behavioral Scientist* 27:211–28.
Bell, D. E. 1982. Regret in decision making under uncertainty. *Operations Research* 30:961–81.
Bishop, R. C., and Heberlein, T. A. 1979. Measuring values of extra-market goods: Are indirect measures biased? *American Journal of Agricultural Economics* 61:926–30.
Chew, S. H. 1983. A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica* 51:1065–92.
Chew, S. H., and MacCrimmon, K. 1979. Alpha utility theory, lottery composition, and the Allais paradox. Working Paper no. 686. Vancouver: University of British Columbia.
Clark, H. H., and Clark, E. V. 1977. *Psychology and Language*. New York: Harcourt Brace Jovanovich.
Einhorn, H. J., and Hogarth, R. M. 1978. Confidence in judgment: Persistence of the illusion of validity. *Psychological Review* 85:395–416.
Ellsberg, D. 1961. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* 75:643–69.
Eraker, S. E., and Sox, H. C. 1981. Assessment of patients' preferences for therapeutic outcomes. *Medical Decision Making* 1:29–39.
Fischhoff, B. 1983. Predicting frames. *Journal of Experimental Psychology: Learning, Memory and Cognition* 9:103–16.
Fishburn, P. C. 1982. Nontransitive measurable utility. *Journal of Mathematical Psychology* 26:31–67.
Fishburn, P. C. 1983. Transitive measurable utility. *Journal of Economic Theory* 31:293–317.
Fishburn, P. C. 1984. SSB utility theory and decision making under uncertainty. *Mathematical Social Sciences* 8:253–85.
Fishburn, P. C. 1985. Uncertainty aversion and separated effects in decision making under uncertainty. Working paper. Murray Hill, N.J.: AT & T Bell Labs.
Fishburn, P. C., and Kochenberger, G. A. 1979. Two-piece von Neumann–Morgenstern utility functions. *Decision Sciences* 10:503–18.
Grether, D. M. 1980. Bayes rule as a descriptive model: The representativeness heuristic. *Quarterly Journal of Economics* 95:537–57.
Grether, D. M., and Plott, C. R. 1979. Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69:623–38.
Hagen, O. 1979. Towards a positive theory of preferences under risk. In M. Allais and O. Hagen (eds.), *Expected Utility Hypotheses and the Allais Paradox*. Dordrecht: Reidel.
Haltiwanger, J., and Waldman, M. 1985. Rational expectations and the limits of rationality: An analysis of heterogeneity. *American Economic Review* 75:326–40.
Hammond, P. 1985. Consequential behavior in decision trees and expected utility. Institute for Mathematical Studies in the Social Sciences Working Paper no. 112. Stanford, Calif.: Stanford University.
Hansson, B. 1975. The appropriateness of the expected utility model. *Erkenntnis* 9:175–93.
Hausch, D. B.; Ziemba, W. T.; and Rubenstein, M. E. 1981. Efficiency of the market for racetrack betting. *Management Science* 27:1435–52.
Hershey, J. C., and Schoemaker, P. J. H. 1980. Risk taking and problem context in the domain of losses: An expected utility analysis. *Journal of Risk and Insurance* 47:111–32.
Kahneman, D.; Knetsch, J. L.; and Thaler, R. H. In this issue. Fairness and the assumptions of economics.
Kahneman, D.; Knetsch, J. L.; and Thaler, R. In press. Perceptions of fairness: Entitlements in the market. *American Economic Review*.
Kahneman, D., and Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–91.

Kahneman, D., and Tversky, A. 1982. The psychology of preferences. *Scientific American* 246:160–73.
Kahneman, D., and Tversky, A. 1984. Choices, values, and frames. *American Psychologist* 39:341–50.
Knetsch, J. L., and Sinden, J. A. 1984. Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *Quarterly Journal of Economics* 99:507–21.
Knez, P.; Smith, V. L.; and Williams, A. W. 1985. Individual rationality, market rationality and value estimation. *American Economic Review: Papers and Proceedings* 75:397–402.
Loomes, G., and Sugden, R. 1982. Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal* 92:805–24.
Luce, R. D., and Krantz, D. H. 1971. Conditional expected utility. *Econometrica* 39:253–71.
Luce, R. D., and Narens, L. 1985. Classification of concatenation measurement structures according to scale type. *Journal of Mathematical Psychology* 29:1–72.
Machina, M. J. 1982. "Expected utility" analysis without the independence axiom. *Econometrica* 50:277–323.
McNeil, B. J.; Pauker, S. G.; Sox, H. C., Jr.; and Tversky, A. 1982. On the elicitation of preferences for alternative therapies. *New England Journal of Medicine* 306:1259–62.
March, J. G. 1978. Bounded rationality, ambiguity, and the engineering of choice. *Bell Journal of Economics* 9:587–608.
Markowitz, H. 1952. The utility of wealth. *Journal of Political Economy* 60:151–58.
Nelson, R. R., and Winter, S. G. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, Mass.: Harvard University Press.
Payne, J. W.; Laughhunn, D. J.; and Crum, R. 1980. Translation of gambles and aspiration level effects in risky choice behavior. *Management Science* 26:1039–60.
Quiggin, J. 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3:323–43.
Raiffa, H. 1968. *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading, Mass.: Addison-Wesley.
Russell, T., and Thaler, R. 1985. The relevance of quasi-rationality in competitive markets. *American Economic Review* 75:1071–82.
Savage, L. J. 1954. *The Foundations of Statistics*. New York: Wiley.
Schelling, T. C. 1981. Economic reasoning and the ethics of policy. *Public Interest* 63:37–61.
Schmeidler, D. 1984. Subjective probability and expected utility without additivity. Preprint Series no. 84. Minneapolis: University of Minnesota, Institute for Mathematics and Its Applications.
Schumpeter, J. A. 1954. *History of Economic Analysis*. New York: Oxford University Press.
Segal, U. 1984. Nonlinear decision weights with the independence axiom. Working Paper in Economics no. 353. Los Angeles: University of California, Los Angeles.
Simon, H. A. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69:99–118.
Simon, H. A. 1978. Rationality as process and as product of thought. *American Economic Review: Papers and Proceedings* 68:1–16.
Slovic, P.; Fischhoff, B.; and Lichtenstein, S. 1982. Response mode, framing, and information processing effects in risk assessment. In R. M. Hogarth (ed.), *New Directions for Methodology of Social and Behavioral Science: Question Framing and Response Consistency*. San Francisco: Jossey-Bass.
Slovic, P., and Lichtenstein, S. 1983. Preference reversals: A broader perspective. *American Economic Review* 73:596–605.
Slovic, P., and Tversky, A. 1974. Who accepts Savage's axiom? *Behavioral Science* 19:368–73.
Smith, V. L. 1985. Experimental economics: Reply. *American Economic Review* 75:265–72.
Stigler, G. J., and Kindahl, J. K. 1970. *The Behavior of Industrial Prices*. New York: National Bureau of Economic Research.

Thaler, R. H. 1980. Towards a positive theory of consumer choice. *Journal of Economic Behavior and Organization* 1:39–60.
Tversky, A. 1969. Intransitivity of preferences. *Psychological Review* 76:105–10.
Tversky, A. 1977. On the elicitation of preferences: Descriptive and prescriptive considerations. In D. E. Bell, R. L. Keeney, and H. Raiffa (eds.), *Conflicting Objectives in Decisions*. New York: Wiley.
Tversky, A., and Kahneman, D. 1981. The framing of decisions and the psychology of choice. *Science* 211:453–58.
Tversky, A., and Kahneman, D. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90:293–315.
von Neumann, J., and Morgenstern, O. 1944. *Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.
Weber, R. J. 1982. The Allais paradox, Dutch auctions, and alpha-utility theory. Working paper. Evanston, Ill.: Northwestern University.
Yaari, M. E. 1984. Risk aversion without decreasing marginal utility. Report Series in Theoretical Economics. London: London School of Economics.