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.

In: The Nobel Prizes 2002, Editor Tore Frängsmyr, [Nobel Foundation], Stockholm, 2003, pp. 449-489

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-

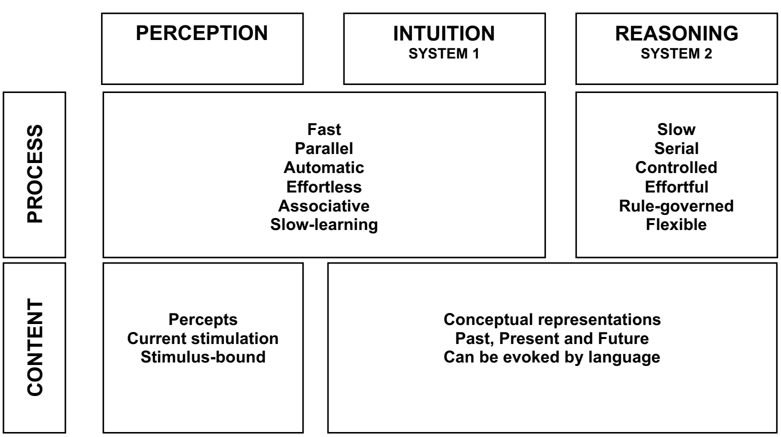

Figure 1.

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

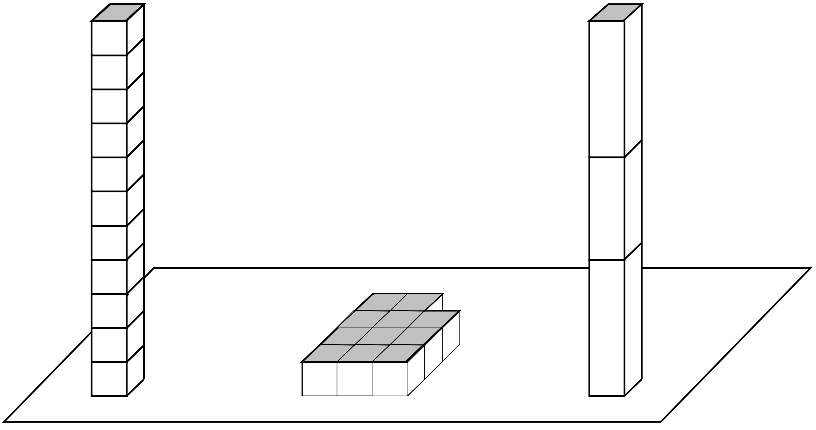

Figure 2a. *Figure 2b.* *Figure 2c.*

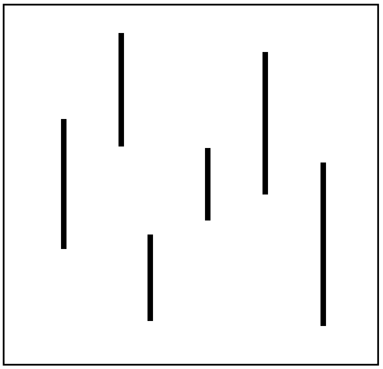

Figure 3.

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).

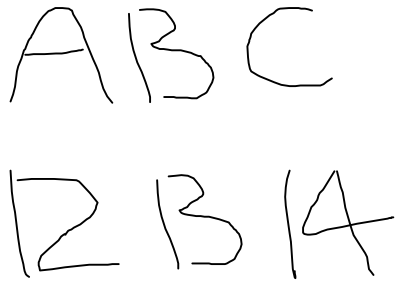

Figure 4.

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

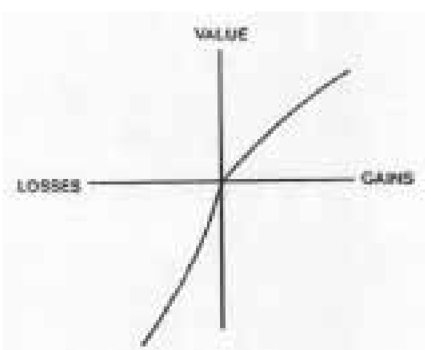

Figure 6.

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

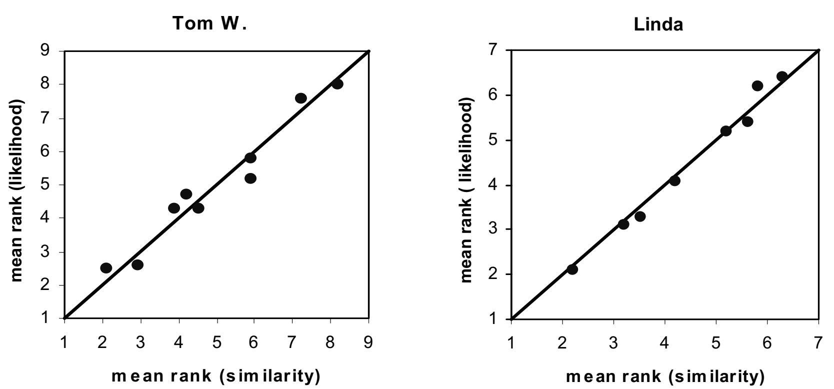

Figure 8.

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).

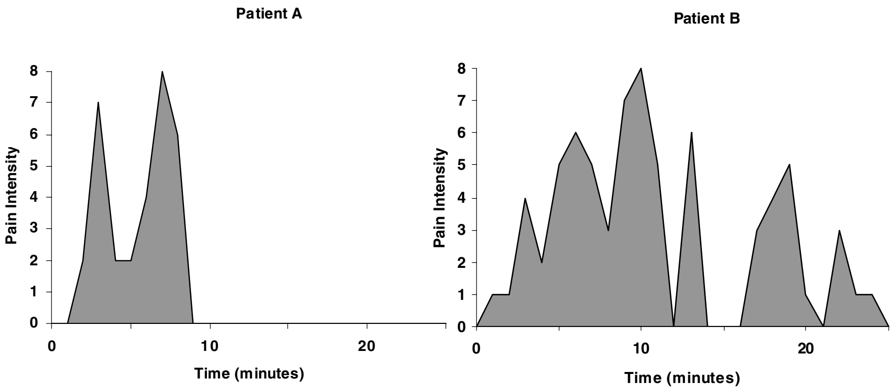

Figure 9. Pain intensity reported by two colonoscopy patients.

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.

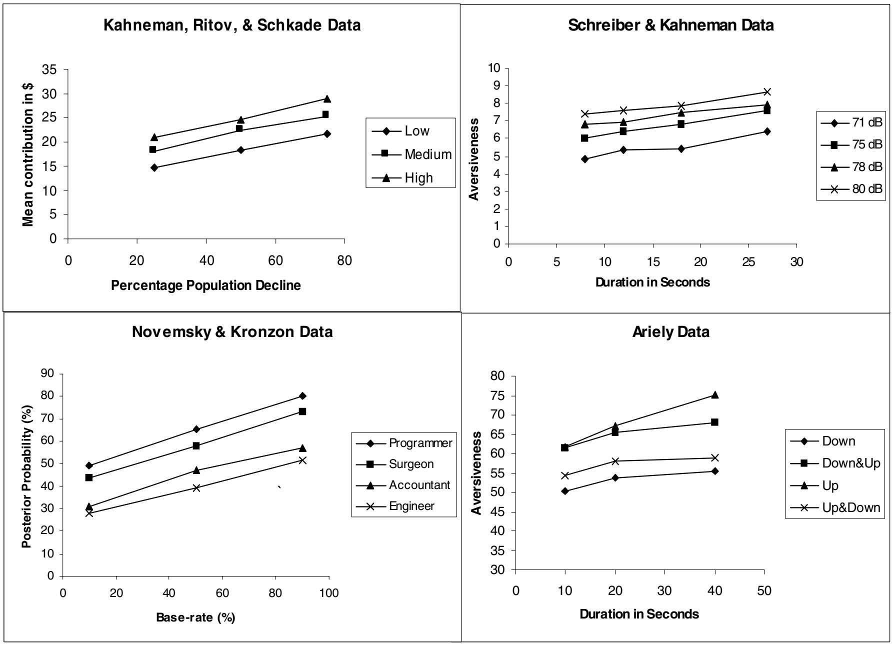

Figure 10. (a) Willingness to pay to restore damage to species that differ in popularity as a function of the damage they have suffered (from Kahneman, Ritov, & Schkade, 1999); (b) Global evaluations of aversive sounds of different loudness as a function of duration for subjects selected for their high 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.