18 What Have We Learned from our Mistakes?

Barbara Mellers & Connson Locke

ABSTRACT

The authors discuss the general steps involved in good decision making and use those steps to organize results from behavioral decision research, illustrating some of the challenges in making important choices. Framing effects, self serving biases, and context effects are a few of the errors and biases presented. The authors also discuss a variety of techniques for reducing biases and errors. The conclusion briefly outlines critiques of the biases and heuristics literature and provides examples of human cognitive strengths, while emphasizing the importance of learning from our mistakes.

CONTENTS

Good Decision Making

To a few lucky people, good decisions come naturally. But to most of us, decisions are difficult, grueling, and even painful. The process requires us to make delicate tradeoffs, sort through complex scenarios, hunt for good ideas, estimate the odds of future states, and answer a voice inside that keeps asking, “Is this what I *really* want?”

Many scholars describe good decision making as a series of interrelated steps. Although there are many ways to break up the process, most researchers agree that the process includes the following stages:

Define the Problem and Set the Goals

The best way to get what you want is to decide what that is. This step could be extremely easy or extremely difficult depending on the problem. A good decision maker asks, "What do I want to achieve? What are my goals and objectives? How will I know if I am successful?"

Gather Information and Identify Options

Important choices require a careful and unbiased search for evidence, and normative theory says that the search for information should continue until the costs outweigh the benefits. New information leads to the discovery of new options which often means the gathering of more information and possibly more options. The best way to find a good option is to identify lots of options. Most problems have more than one solution, but the solutions may not be obvious. Creating, finding, devising, or isolating alternatives is a key part of decision making.

Evaluate the Information and the Options

Here, the decision maker should ask, “What really matters to me? Why do I care?” Answers to these questions should lead to the evaluation of outcomes and the measurement of beliefs. Then, the decision maker combines evaluations and beliefs to form overall assessments of the options.

Make a Choice

Criteria for the selection of an option should be based on the decision maker’s goals and objectives. There may be a single criterion, such as maximizing pleasure, or multiple criteria, such as maximizing profit, minimizing time, and minimizing risk.

Implement the Choice and Monitor the Results.

This step may be the most important of all. All prior steps are useless without a commitment to action. Moreover, choices must be monitored. Since most decisions rarely proceed as planned, a decision maker should keep a watchful eye on the consequences and be prepared to make corrective adjustments as needed. Good decision makers should be committed to their decisions, but flexible in their approaches.

To Error is Human

Those are the steps. How well can people do them? For the last five decades, behavioral decision researchers have been asking that question along with a variety of others, including “What, if anything, do people do instead?” The most well known research program in behavioral decision making was started by Kahneman and Tversky in the early 1970s. Human judgments and decisions were held up to scrutiny and evaluated in light of the standards set by normative theory. Results were surprising, intriguing, and quite perplexing. Kahneman and Tversky proposed that human judgments were not governed by normative theory, but could be described in terms of heuristics and biases.

Initially, there were three heuristics: availability, representativeness, and anchoring and adjustment. The availability heuristic states that people assess the probability of an event based on the degree to which instances come to mind. What comes to mind is based on vividness and recent experience. Ross and Sicoly (1979) illustrated this heuristic when they asked husbands and wives to estimate the extent to which each was responsible for domestic activities, such as cooking, cleaning, and shopping. When summed together, the average couples' percentage was 130%. Each partner could quickly remember instances in which he or she took out the garbage, unloaded the dishwasher, or folded the clothes. The other person's contributions were less accessible. Similar effects were found with a student and faculty member who estimated the percentage of work they did on a joint research project. When summed, the average joint percentage was 130%. Because people have vivid memories of their own efforts, they tend to overweight their contributions relative to the contributions of their partners.

The second heuristic is called representativeness. This heuristic states that, when making a judgment, people consider the degree to which the specific information represents a relevant category or population. The Linda problem (Tversky & Kahneman, 1983) is a compelling example of this heuristic. In the Linda problem, participants are told:

> 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 anti-nuclear demonstrations. Rank the following descriptions in terms of the probability they describe Linda.

1. Linda is active in the feminist movement
2. Linda is a bank teller.
3. Linda is a bank teller and is active in the feminist movement.

The description of Linda is representative of a feminist, not a bank teller. If people make their judgments according to the description of Linda, they will say that the first statement is most likely, followed by the third statement, and finally the second statement. This is exactly what happens, even though this order violates an important rule of probability called the conjunction rule. The conjunction rule states that the combination of two events can never be more probable than the chance of either event by itself. In contrast, the representativeness heuristic says we judge probability in terms of the similarity of the target stimulus to a category. The Linda problem shows that the rules of probability differ from the rules of similarity.

Finally, the third heuristic is called anchoring and adjustment. This heuristic asserts that people make probability judgments by starting with a number that easily comes to mind and adjusting for additional information. But adjustments are often insufficient. Anchors may be valid and useful cues, or they may be completely irrelevant. Effects of irrelevant anchors were demonstrated by Russo and Schoemaker (1989) who asked respondents to write down the last 3 digits in their telephone number. Then they were asked whether Attila the Hun was defeated in Europe before or after the year defined by the 3 digit number. Finally, the respondents were asked to estimate the year of Attila's defeat. Telephone numbers are obviously unrelated to Attila the Hun. Nonetheless, the numbers influenced historical estimates. Estimates of the year of Attila the Hun’s defeat were higher among participants with larger 3 digit numbers and lower among participants with smaller 3 digit numbers. A variety of similar examples are well documented (e.g. Tversky & Kahneman, 1974; Strack & Mussweiler, 1997; Chapman & Johnson, 1994).

Even experts can fall prey to anchoring and adjustment effects. In one study, practicing auditors estimated the incidence of executive fraud. Before providing their estimates, auditors in one condition were asked if the incidence of fraud was more than 10 in 1,000 among companies audited by Big Five accounting firms. In the second condition, auditors were asked if the incidence was more than 200 in 1,000. Auditors then gave their estimates. In the first condition, the average estimate was 16.52 per 1,000, while in the second, the average estimate was 43.11 per 1,000, over twice as large. (Joyce & Biddle, 1981)

Kahneman and Frederick's (2002) more recent view of heuristics and biases is that the first two heuristics – availability and representativeness – can be subsumed under a single heuristic called attribution substitution.[1]

If a target attribute is relatively inaccessible, people substitute it with something that more easily comes to mind. Many attributes may be substituted for the relevant attribute, especially those that are vivid and emotional. An example in which people substitute emotional reactions for monetary values comes from the literature on contingent valuation. Contingent valuation is a method used by economists to assign a monetary value to a public good that would never be bought or sold in the marketplace, such as clean air, clean beaches, or clean lakes. Economists ask participants to state the maximum amount they would be willing to pay to either maintain a public good or restore it to a previous state. Judged values reported in these surveys are not always consistent with common properties of real economic values.

In one study, Desvouges, Johnson, Dunford, Hudson, Wilson, and Boyle (1993) asked different groups of participants to state the maximum amount they would be willing to pay to clean up oil ponds that had led to the deaths of 2,000, 20,000 or 200,000 migratory birds. Average amounts of $80, $78, and $88, respectively, were relatively insensitive to the number of birds that would be saved. Kahneman, Ritov, and Schkade (1999) argued that the death of birds evokes a feeling of outrage, and that emotional response is mapped onto a monetary scale. Similar degrees of outrage are associated with a wide range of economic consequences (e.g., 200 to 200,000 birds), so unlike real economic values, judged values remain constant.

If people use these heuristics and apply them incorrectly, how well do they make decisions? Behavioral decision researchers have much to say about the types of mistakes that can occur when, where, how, and why. We now present some well-known errors and biases in the context of the five steps of good decision making. Our list is by no means exhaustive; it is purely intended to illustrate how natural behavioral tendencies can interfere with good decisions.

Define the Problem and Set the Goals

When people think about choices, they often accept and use information as it was received. This tendency can lead to systematic differences in preference known as framing effects. Framing is a bit like taking a photograph. One must decide how far away to stand, what to include, and which elements define the figure and the ground. Behavioral decision researchers have found that, when the same choice is presented using different frames, people often reverse their preferences.

A classic example is the Asian Disease Problem (Tversky & Kahneman, 1981). Both versions of the problem state:

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

In the gain frame, participants are told:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no one will be saved.

In the loss frame, participants read:

If Program A is adopted, 400 people will die

If Program B is adopted, there is a 1/3 probability that no one will die, and a 2/3 probability that 600 people will die.

Despite the different descriptions, Program A is the same across both frames, and Program B is the same across both frames. Tversky and Kahneman found that 76% preferred Program A in the gain frame, while 71% preferred Program B in the loss frame. Participants' preferences were risk averse with gains and risk seeking with losses.

Frames are subtle, yet consequential. Most consumers would agree that ground beef sounds better if the package says 75% lean than 25% fat (Levin & Gaeth, 1988). A hospital policy sounds more effective if the report indicates that 75% of beds were full than 25% of beds were empty. Even a change in pricing might seem fair if described as a discount, but unfair if described as a surcharge. Northwest was one of the first airlines to charge passengers $10 more for tickets purchased at the airport than they did for tickets purchased online (NYT, August 25, 2004). The headline of the Times article read, "Why Fly? Get Charged $10 Just to Show Up!" Company executives pointed out that JetBlue had the same $10 fee. JetBlue quickly responded by saying that Northwest executives were wrong. JetBlue charged standard fares for tickets purchased at the airport, but offered a $10 discount to customers who bought tickets electronically.

Framing effects can have powerful financial consequences. Johnson, Hershey, Meszaros, and Kunreuther (1993) described framing effects in the insurance industry. In 1988, the standard auto policy in New Jersey did not allow drivers the right to sue for pain and suffering from minor injuries, although they could purchase that right with a higher-priced policy. Only 20% of New Jersey drivers bought the more expensive policy. In 1990, the standard auto policy in Pennsylvania included the right to sue, and 75% of Pennsylvania drivers purchased it. Johnson, et al. (1993) estimated that Pennsylvanians spent $200 million more on auto insurance than they would have if the default had been the cheaper option. Unless decision makers are able to think about a problem from different perspectives, they may be "framed" by information as it appears.

Gather Information and Identify Options

Some say that the greatest danger to good intelligence gathering is not insufficient time, rather the mental biases and distortions that we bring to the search process. Unfortunately, we don't always know what we need to know, and we focus inordinate attention on evidence that confirms our beliefs and hypotheses. This tendency runs directly counter to the scientific method. With the scientific method, we try to disprove—not prove—our hypotheses. But thinking negatively takes additional effort.

A classic example of the confirmation bias comes from a study by Wason (1960). He asked subjects to imagine that the sequence of three numbers, such as “2-4-6”, follows a rule. The task is to discover the underlying rule by generating sequences and receiving feedback about whether those sequences are consistent or inconsistent with the rule. Suppose the real rule is “any three ascending numbers”. When given an initial starting sequence of “2-4-6”, subjects often assume the rule is "numbers that go up by two". They test the hypothesis with sequences such as “1-3-5” or “8-10-12”.

It is fairly unusual for subjects to test the hypothesis with a disconfirming sequence, such as “10-15-20” or “6-4-2”.

Another example is the four-card problem. Subjects are shown four cards with a number on one side and a letter on the other. The cards are labeled, "E", "4", "7" and "K". Subjects are asked to identify only those cards that must be checked to test the claim that "If a card has a vowel on one side, it must have an even number on the other side." Most subjects choose “E” (to see if it has an even number on the other side) and "4" (to see if it has a vowel on the other side). But both of these tests confirm the claim. The correct answer is "E" and "7". If the claim is true, "E" should have an even number on the other side, and "7" should have a consonant on the other side. The other two cards are irrelevant to the claim.

The confirmation bias, or the tendency to gather information that supports our hypotheses, has been attributed, at least in part, to self-serving attributions. Psychologists have identified a variety of ways in which people maintain positive views of themselves. Overconfidence is the tendency to be more confident in one’s own ability than reality dictates. In a typical overconfidence experiment, participants are given a series of true-false questions, such as “The population of London is greater than that of Paris.” They answer “True” or “False” and then judge their confidence that they are correct. If they were completely unsure of their answer, they would say “50%”. If they were absolutely sure they were correct, they would say “100%”. Average confidence ratings over questions are significantly greater than the actual percentage of correct items (Fischhoff, Slovic, & Lichtenstein, 1986). In fact, when participants say they are 100% confident, their accuracy rates for those items are typically around 75%.

Overconfidence goes far beyond true-false questions on general knowledge tests. It has been observed in physicians (Lusted, 1977), clinical psychologists (Oskamp, 1965), lawyers (Wagenaar &

Keren, 1986), negotiators (Neale & Bazerman, 1992), engineers (Kidd, 1970), security analysts (Stael von Holstein, 1972) and eyewitness testimonies (Sporer, Penrod, Read, & Cutler, 1995). Confidence in one's accuracy guarantees very little, even among professionals.

Closely related to overconfidence is the "above average" effect. People perceive themselves as being better than others on most desirable attributes, including honesty, cooperativeness, health, intelligence, managerial skill, and even driving ability (Babcock & Loewenstein, 1997; Larwood & Whittaker, 1977). A survey conducted with 1 million high school students in 1976-77 by the College Board found that 70% of students rated themselves as above average in leadership ability, and only 2% rated themselves as below average.

People also seem to have undue optimism about their future. Predictions are systematically better and brighter than reality. For example, MBAs overestimate the number of offers they will receive, the magnitude of their starting salary, and how early they will get their first offer (Hoch, 1985). Newlyweds almost uniformly expect their marriages to last a lifetime (Baker & Emery, 1993), and financial investors predict that, unlike others, their forecasts will beat the market averages (Barber & Odean, 2000).

Self-serving biases can even influence perceptions of fairness in distributions of scarce resources. In an illustrative experiment, Messick and Sentis (1983) told participants to assume they had worked with another person on the same job. Their task was to allocate a sum of money between themselves and the other person as they saw fit. In some conditions, participants learned that they (or the other person) had either worked twice as long or had accomplished twice as much. Most participants gave themselves more than half if they worked longer or harder. But if the other person worked longer or harder, participants thought that an even split was fair. Apparently, it is hard to view one's efforts and accomplishments from a neutral point of view.

Loewenstein, Issacharoff, Camerer, and Babcock (1993) also illustrated self-serving biases in fairness in a negotiation settlement. They randomly assigned participants to the role of the defendant or the plaintiff and gave them identical case materials about a legal dispute. The task was to negotiate a monetary amount that the defendant would pay the plaintiff, and both parties were financially penalized if they failed to reach an agreement. If an agreement was not reached, a judge would determine a fair amount for the defendant to pay the plaintiff. Before negotiating, participants were asked to predict the judge's settlement. Plaintiffs' predictions of the judge's settlement were substantially higher than defendants' predictions. Perceptions of "fair" settlements were self-serving perceptions, despite incentives for accuracy.

If we are overly confident about our abilities, convinced that we are better than others, and excessively optimistic about our futures, we won't need to gather as much information when making decisions. Furthermore, if the information we gather confirms our hunches, we should expect smooth sailing, shouldn't we?

Evaluate the Information and the Options

Unlike physics, where the assessments of length and weight do not depend on the stimuli previously measured, psychological assessments vary with the context. Responses to the same stimulus differ depending on the surrounding stimuli. For example, a 10 pound sack of apples seems light after carrying a 50 pound child, but heavy after carrying a soft drink.

Stimulus contexts can be local or global. In the domain of choice, the local context refers to the particular choice set. A classic local contextual effect is called asymmetric dominance (Huber, Payne, & Puto, 1982). Participants choose between two options, A and B, each of which is described by two attributes, such as price and quality. Later, participants are given the same two options in a choice set that contains a third option, C, that is dominated by B and asymmetrically dominated (on one attribute) by A. With C in the choice set, the relative preference for B increases. According to most theories of choice, preferences should never increase as more options are added to the consideration set. But that is exactly what happens.

The global context, again in the domain of choice, refers to all of the choice sets. Mellers and Cooke (1996) found that preferences for a given choice can reverse when placed in different global contexts. They asked students to make choices between pairs of apartments described in terms of monthly rent and distance to campus. In one context, rents went from $100 to $1000, and distances varied from 10 to 26 minutes. In another, rents went from $100 to $400, and distances varied from 10 to 50 minutes. Both contexts included a choice between a $400 apartment that was10 minutes from campus and a $200 apartment that was 26 minutes away. When the context included more expensive apartments, 60% of students preferred the $400 apartment that was 10 minutes away. When the context included greater distances, 64% preferred the $200 apartment that was 26 minutes away. The less desirable attribute did not seem as bad when the context included apartments that were worse.

Another global context effect is the order in which questions are asked. Order effects are changes in response due to sequence. Strack, Martin, and Schwarz (1988) demonstrated an amusing order effect by asking college students two questions: “How happy are you with your life in general?” and “How many dates did you have last month?” The correlation between students’ responses was negligible when the happiness question came first. But that correlation rose to .66 when the dating question came first. The dating question evoked a general feeling of satisfaction or dissatisfaction with one’s life, and that feeling was used to answer the happiness question.

Primacy and recency effects are order effects in which people place disproportionate weight on the initial stimulus or final stimulus, respectively. Lawyers scheduling the appearance of witnesses for court testimony and managers scheduling speakers at meetings take advantage of these effects when they organize sequences of events. Primacy and recency effects have been demonstrated in numerous studies of memory. Primacy effects have also been found in impression formation. Participants are asked to form an impression of a person who has been described with various traits (Asch, 1946). Half of the participants are told the person is envious, stubborn, critical, impulsive, industrious, and intelligent. The other half are told the person is intelligent, industrious, impulsive, critical, stubborn, and envious. Although the two lists contain identical traits, participants describe the person as less likeable when the negative traits come first. Over-reliance on first impressions is a form of primacy. Recency effects occur in debates when the person who gets the "final word" seems to win. The last word is often easier to remember than words uttered earlier.

If we allow contextual effects and order effects to distort our evaluation of the options, then we might not make the best decision. But these are not the only challenges we face when making choices.

Make a Choice

Several problems can occur at this step. People may ignore options, outcomes, information, or values. They may engage in wishful thinking and tell themselves that an unfortunate event will not happen to them. They may be lazy and satisfied by selecting the first option that seems to be "good enough", even though a far better option might lie just around the corner (Simon, 1956). People also base their choices on immediate emotional reactions that might not reflect their true values (Slovic, Finucane, Peters, & MacGregor, 2004).

During the choice process—especially with difficult choices—people often stick with the status quo, possibly because the perceived costs of a change seem larger than the perceived benefits. Samuelson and Zeckhauser (1988) demonstrated the status quo effect in a study of decisions about health plans made by Harvard employees. In 1980, Harvard employees could choose one of four health plans. By 1986, that number had increased to eight. Old enrollees who had made their choices before 1980 kept the status quo more often than new enrollees who came between 1980 and 1985. The low rate of transfer among those who had been with their plans longer was consistent with the status quo bias.

Status quo biases can develop remarkably fast. In one experiment, Knetsch (1989) gave students a university coffee mug and then asked them to complete a short questionnaire. Students had the mug in their possession while they worked on the questionnaire. When they were finished, students were offered a choice between keeping their mug or exchanging it for a large, Swiss chocolate bar. Only 11% wanted to trade. Knetsch (1989) then repeated the experiment but the initial gift was the Swiss chocolate bar. After completing the questionnaire, students could either keep their Swiss chocolate bar or exchange it for a mug. Only 10% wanted to trade.

The status quo bias can have serious consequences. Ritov and Baron (1990) asked participants to imagine that their child had a 10 out of 10,000 chance of death from a flu epidemic. A vaccine could prevent the flu, but for some, it was fatal. Participants were asked to indicate the highest risk they would accept for the vaccine, before they would refuse it. Most subjects answered that the chance of death from the vaccine had to be well below 9 out of 10,000. Anyone who insisted the vaccine be safer than 10 out of 10,000 was putting their child at greater risk by refusing the vaccination. In follow up questions, participants confessed that they would feel greater regret if the child died from the vaccine than from the flu. The sense of regret is another driving force behind the status quo bias.

Some people can't free themselves from the status quo; others can't help choosing an option into which they already invested time or money, even if it is not the best option. This finding is known as the sunk cost effect. Most people find it hard to leave a meal unfinished at a restaurant or walk out of a bad movie. They want to "get their money's worth". Nonetheless, past costs should not influence choices; only future costs should matter.

An example of sunk cost effects comes from the health club industry. Gourville and Soman (1998) obtained usage data from members of a club that charged dues twice a year. Attendance at the health club was highest in the months when dues were paid, then declined over the next 5 months, only to jump again when the next dues were paid.

Thaler (1999) says that although sunk costs linger, they do not hang on indefinitely. A thought experiment illustrates the point. Suppose you buy a pair of shoes. They feel very comfortable in the store, but the first day you wear them, they hurt badly. A few days later you try them again, but they hurt even worse than they did the first time. What happens next? The more you paid for the shoes, the more times you will try to wear them. This may be rational, especially if you need to replace the shoes with another pair. Eventually you stop wearing the shoes, but you do not throw (or give) them away. The more you paid for the shoes, the longer they sit in the back of your closet. Finally, you throw away (or give away) the shoes. This is when the payment has been fully depreciated.

Status quo biases and sunk cost effects are potential pitfalls whenever we make choices. These mistakes suggest that we are more likely to select options with which we are familiar or into which we have already invested time or money.

Implement the Choice and Monitor the Results

Escalation of commitment and the hindsight bias are two important errors that can occur in this step. Recall that the sunk cost effect occurs when decision makers choose an option because of a prior investment, even when it is not the best option (e.g., Arkes and Blumer 1985). Escalation of commitment (Staw, 1976) is the tendency to continue investing in that suboptimal option. Decision makers are more likely to throw good money after bad when 1) they are personally responsible for the initial investment, 2) the initial investment is failing, and 3) they have publicly announced their position.

In organizational settings, it is not easy to decide when to fire a slow employee, drop a disappointing new product, or loan more money to a company that lost the initial loan. De-escalation might only happen when and if the person who made the initial decision leaves the organization. In a longitudinal study of California banks, Staw, Barsade, and Koput (1997) found that turnover rates among managers was associated with de-escalation of commitment to bad loans, whereas turnover rates among the board of directors was not. Managers appeared to be justifying their decisions by escalating their commitments, and de-escalation was possible only after a change in management.

The hindsight bias is the tendency to misremember the degree to which one accurately forecasted an outcome. We remember being more accurate than we actually were. When looking back and thinking about an initial forecast, we find it difficult to ignore the outcome that occurred. In addition, we are better able to recall information that was consistent rather than inconsistent with the outcome (Dellarosa & Bourne, 1984). Some call the hindsight bias the "knew-it-all-along" effect because it leads people to view whatever occurred as relatively inevitable (Fischhoff, 1975). In an early study of the hindsight bias, Fischhoff and Beyth (1975) asked students to predict what would happen in 1972 when Nixon visited China and the former Soviet Union. Would Nixon meet with Mao? Would the US and the former Soviet Union form a joint space program? Two weeks later, the experimenters asked the students to recall their earlier probability estimates. 67% of students remembered themselves as being more accurate than they actually were. As time went on, the bias increased, with 84% showing the hindsight bias after 3 to 6 months.

Hindsight biases have been documented in elections (Leary, 1982; Synodinos, 1986), medical diagnoses (Arkes, Saville, Wortmann, & Harkness, 1981), business ventures (Bukszar & Connolly, 1988) and historical records (Fischhoff, 1975). To evaluate past decisions and thereby improve future decisions, we must put ourselves into the mindset of the decision maker prior to outcome feedback. But research shows it isn't easy to make that cognitive leap.

Minimizing Mistakes

As described above, each step in the process of good decision making is vulnerable to biases and errors. Researchers have explored ways of reducing or eliminating these mistakes. One approach targets the information used by the decision maker. Combining frames, using frequency formats, or disclosing conflicts of interest change the information with which a decision maker is presented. Another approach targets the decision maker's thought processes. Increasing accountability, providing incentives, and training people to reduce overconfidence change decision makers' motivation and ability to calibrate. In this section, we discuss the strengths and weaknesses of these methods.

Combining Frames

It makes intuitive sense that framing effects could be reduced if decision makers approached a problem from multiple perspectives. But there are different ways to consider multiple perspectives.

For example, multiple frames could be presented sequentially with the same choice nested in different frames. If preferences reversed, those shifts could be explored and resolved. Research that has examined this form of multiple frames typically gives subjects a choice using one frame. After some unrelated tasks, subjects are given the same choice in another frame (LeBoeuf & Shafir, 2003; Frisch, 1993; Levin et al., 2002; Stanovich & West, 1998). Although some people have consistent preferences across frames, many others change their preferences. Decision analysts would need to discuss the frames in more detail with these people.

Another way to present multiple frames is to combine them into a single choice. For example, the Asian Disease problem could be expressed as:

If Program A is adopted, 200 people will be saved and 400 people will die.

If Program B is adopted, there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that 600 people will die.

This combined format does not permit decision makers the "luxury" of preference reversals. They must confront any conflicting goals or competing desires directly in order to make their choice.

To examine the combined format, we administered different versions of a questionnaire to three different groups of undergraduates at the University of California Berkeley, with between 162 and 187 participants in each group. Each version included three framing problems: the Asian Disease problem from Tversky and Kahneman (1981), a simplified version of the Layoff problem from Messick and Bazerman (1996), and the Medical problem from McNeil et al. (1982).

The gain and loss frames of the Asian Disease problem replicated Tversky and Kahneman's earlier results. The majority of participants preferred Program A, the safer option, in the gain frame (76%), and Program B, the riskier option, in the loss frame (71%). That is, preferences were risk averse in the gain domain and risk seeking in the loss domain. In our examination of the combined frame, where gains and losses are both vivid, the majority of respondents preferred Program A (58%), the risk averse option. The combined frame more closely resembled the gain frame than the loss frame.

Next we examined the Layoff problem. All three versions said,

A large car manufacturer has recently been hit with a number of economic difficulties, and it appears as if 9,000 employees must be laid off. The vice president of production has been exploring alternative ways to avoid the crisis. She has developed two plans.

In the gain frame, participants were told:

Plan A will result in the saving of 3,000 jobs.

Plan B has a 1/3 chance of saving 9,000 jobs and a 2/3 chance of saving no jobs.

In the loss frame, participants read:

Plan A will result in the loss of 6,000 jobs.

Plan B has a 1/3 chance of losing no jobs and a 2/3 chance of losing 9,000 jobs.

Finally, the combined frame stated that:

Plan A will result in the saving of 3,000 jobs and the loss of 6,000 jobs.

Plan B has a 1/3 chance of saving 9,000 jobs and a 2/3 chance of losing 9,000 jobs.

The majority of respondents preferred Plan A in the gain frame (71%) and Plan B in the loss frame (63%). Once again, the combined frame revealed a tendency toward risk aversion with most participants choosing Plan A, the safer option (73%).

Then we examined the Medical problem from McNeil et al. (1982) in which subjects chose between two types of treatment for lung cancer: surgery or radiation. Data given to subjects were presented in terms of survival rates or mortality rates. Surgery offers a better long term life expectancy, but a 10% chance of immediate death on the operating table. For this reason, surgery appears more desirable in the survival frame than the mortality frame. Both versions of the problem stated:

> Surgery for lung cancer involves an operation on the lungs. Most patients are in the hospital for two or three weeks and have some pain around their incisions; they spend a month or so recuperating at home. After that, they generally feel fine. Radiation therapy for lung cancer involves the use of radiation to kill the tumor and requires coming to the hospital about four times a week for six weeks. Each treatment takes a few minutes and during the treatment, patients lie on a table as if they were having an X-ray. During the course of the treatments, some patients develop nausea and vomiting, but by the end of the six weeks they also generally feel fine. Thus, after the initial six or so weeks, patients treated with either surgery or radiation therapy felt the same. Now imagine you are a patient given the choice between these two types of treatment. You are given the following statistics regarding the consequences of each type of treatment:
>
> In the survival frame, participants were told:
>
> Of 100 people having surgery, 90 will live through the surgery, 68 will be alive at the end of the first year, and 34 will be alive at the end of 5 years.
>
> Of 100 people having radiation therapy, all live through the treatment, 77 will be alive at the end of the first year, and 22 will be alive at the end of 5 years.
>
> In the mortality frame, they were told:
>
> Of 100 people having surgery, 10 will die during the treatment, 32 will have died by 1 year, and 66 will have died by 5 years.
>
> Of 100 people having radiation therapy, none will die during treatment, 23 will die by 1 year, and 78 will die by 5 years.

McNeil et al. found an overall preference for surgery in both frames. But the relative preference for surgery was stronger in the survival frame (82%) than in the mortality frame (53%). McNeil et al. also explored preferences in a combined frame which read:

> Of 100 people having surgery, 10 will die during treatment and 90 will live through the treatment. A total of 32 people will have died by the end of the first year and 68 people will be alive at the end of the first year. A total of 66 people will have died by the end of 5 years and 34 people will be alive at the end of 5 years.
>
> Of 100 people having radiation therapy, none will die during the treatment (i.e., all will live through the treatment). A total of 23 will have died by the end of the first year and 77 will be alive at the end of the first year. A total of 78 people will have died by the end of 5 years and 22 people will be alive at the end of 5 years.

They found that 60% preferred surgery. The combined frame resembled the mortality frame more than the survival frame.

Our results showed an overall reduction in the preference for surgery, but the same relative preference shift. More participants chose surgery in the survival frame (54%) than in the mortality frame (32%). In our combined frame, 57% of participants preferred surgery. Preferences in the combined frame mimicked the survival frame more than the mortality frame.

In a similar study that also examined separate and combined frames for the Medical problem, Armstrong, Schwartz, Fitzgerald, Putt, and Ubel (2002) obtained results that resembled ours. Preferences for surgery were reduced relative to those found by McNeil et al., and preferences in the combined frame resembled those in the survival frame. Furthermore, Armstrong et al. showed that participants understood the information better in the combined frame and the survival frame than in the mortality frame. Their data provide further evidence that the combined frame better reflects preferences.

We do not mean to imply that all combined frames will lead to risk averse preferences, although the regularity of the results is intriguing. We do, however, wish to point out that combined frames are a useful way of presenting multiple reference points and possibly assessing more stable and robust preferences.

Frequency Formats

Probabilistic inferences are often hard to make, even for those with statistical training. Some scholars have asked whether understanding and accuracy of inferences could be improved if the information were presented in another format. Frequencies seem to be easier to work with than probabilities. Hoffrage and Gigerenzer (1998) asked a group of physicians to make inferences about the presence of a disease given a positive test result on a mammogram. Half of the physicians were given the following question in a probability format which read:

> The probability of breast cancer is 1%. The probability of a positive test given breast cancer is 80%, and the probability of a positive test given no breast cancer is 10%. What is the probability that a woman who tests positive for breast cancer actually has breast cancer?

The other physicians were given the same question using a frequency format as follows:

> 10 of every 1,000 women have breast cancer. 8 of those 10 women with breast cancer will test positive, and 99 of the 990 women without breast cancer will also test positive. How many of those women who test positive actually have breast cancer?

The correct answer is 7.5% (or 8 out of 107 women). Hoffrage and Gigerenzer found that only 8% of physicians made the correct inference in the probability format, but almost half (46%) made the correct inference in the frequency format. Frequency formats improve Bayesian reasoning in both laypersons (Gigerenzer & Hoffrage, 1995) and experts (Hoffrage & Gigerenzer, 1998).

Frequency formats can also reduce conjunction errors. Tversky and Kahneman (1983) presented participants with the following question in a probability format:

> A health survey was conducted in a sample of adult males in British Columbia, of all ages and occupations. Please give your best estimate of the following values:
>
> What percentage of the men surveyed have had one or more heart attacks?
>
> What percentage of the men surveyed are both over 55 years old and have had one or more heart attacks?

Only 35 % avoided the conjunction fallacy by assigning a higher percentage to the first question. Participants given the question in a frequency format read:

> A health survey was conducted in a sample of 100 adult males in British Columbia, of all ages and occupations. Please give your best estimate of the following values:
>
> How many of the 100 participants have had one or more heart attacks?
>
> How many of the 100 participants are both over 55 years old and have had one or more heart attacks?

This time, 75% assigned a higher frequency to the first question than the second, consistent with the conjunction rule.

Evolutionary psychologists have argued that success in our ancestral environment required an understanding of frequencies, but not probabilities. Therefore, according to these psychologists, it should not be surprising that inferences are more accurate when information is provided with frequencies than with probabilities. Frequency formats do not fix all probability errors (Mellers, Hertwig, & Kahneman, 2001), but they do fix some. For an overview of when frequency formats help and when they do not, see Sloman, Over, Slovak, and Stibel (2003).

Disclosure

Conflicts of interest occur when personal interests clash with professional interests in financial, medical, legal, or scientific domains. A common procedure for constraining self interest is disclosure. Medical journals require researchers to disclose the sources of their research funding. Stock analysts are required to disclose their financial conflicts of interest when they make public stock recommendations. The McCain-Feingold Act mandates public disclosure of political contributions. Even warning labels on products, such as cigarettes and alcohol, are designed to give consumers full information about the consequences of consumption.

The widespread use of disclosure is based on the assumption that revealing a conflict of interest will either result in the advisor providing less biased advice or the decision maker discounting the advisor's advice, or both, thus allowing an unbiased decision, despite the conflict of interest. A recent study by Cain, Loewenstein, and Moore (2005) suggests that disclosure might not only fail to solve the problems created by conflict of interest but may sometimes make matters worse. In their experiment, subjects were randomly assigned the role of an advisor or an estimator. The estimators' task was to guess the total value of a jar filled with coins; they were given 10 seconds to view the jar from a distance of 3 feet. Advisors were allowed more time and better viewing. Advisors gave their recommendations to estimators before estimators made their guesses. Estimators were always paid based on their accuracy, but advisor payments varied across conditions. In the control condition with no conflict of interest, advisors were paid if estimators guessed correctly. In the condition with conflict of interest and no disclosure, advisors were paid if estimators overvalued the jar, and estimators were not aware of the advisors' incentives. In the condition with conflict of interest and disclosure, advisors were paid if estimators overvalued the jar, and estimators *were* aware of the advisors' incentives.

Estimates were made of several jars, and the average value was $18.16. With no conflict of interest, the average advisor estimate was $16.48. With conflict of interest and no disclosure, the average advisor estimate was $20.16. Did disclosure make the advisors give more accurate estimates? No, the average advisor estimate was $24.16, even higher with disclosure. Disclosure increased advisor bias, but it also increased estimators' discounting. With disclosure, estimators discounted the recommendations by an average of $2.50 more compared to the non-disclosure condition. This discounting did not correct for the increased bias in advisor recommendations of $4. Estimators earned the least amount of money in this condition.

Advisors may feel that, if they disclose their self interests, they are less liable to recipients. Decision makers who learn that an advisor has a conflict of interest should rely more heavily on their private information, but it is difficult to determine how much the conflict of interest has affected the advisor. And, in a real life situation, if decision makers are naturally trusting of their advisors and unsure of their own information, they may interpret the advisor's disclosure as a reason to believe the advisor's information is even more credible.

Accountability

Common sense suggests that people will be more careful in their judgments and decisions if they expect to be called on to justify their choices or actions to others. Research shows that accountability can be helpful in some circumstances (Lerner & Tetlock, 1999). Decision makers are often able to view the problem from different perspectives when anticipating the objections of an audience with unknown views (Ashton, 1992; Weldon & Gargano, 1988; Tetlock et al., 1989). Accountability can also be helpful when decision makers realize they will be accountable to an audience before, rather than after, they examine the evidence (Tetlock, 1983). Predecisional accountability to an unknown audience can reduce cognitive biases.

What, exactly, does accountability do? Predecisional accountability to an unknown audience makes people less sensitive to the order of information (i.e., primacy and recency effects) (Kruglanski & Freund, 1983; Tetlock, 1983), less overconfident (Tetlock & Kim, 1987; Seigel-Jacobs & Yates, 1996), and less susceptible to conjunction errors (Simonson & Nye, 1992). Accountability also makes people less likely to draw inferences from incomplete evidence, more willing to revise their opinions in response to unexpected feedback, and less likely to be swayed by vivid, but irrelevant, information (Lerner & Tetlock, 1999).

Unfortunately, that's not all. With predecisional accountability to an unknown audience, people are more sensitive to status quo effects (Samuelson & Zeckhauser, 1988), compromise effects or the tendency to prefer the "middle" option (Simonson & Nye, 1992), asymmetric dominance or the tendency to choose the option that dominates another option on one dimension (Simonson, 1989), and dilution effects or the tendency to use nondiagnostic information (Tetlock & Boettger, 1989). Predecisional accountability to an unknown audience can reduce some biases, especially those that result from too little effort or a lack of understanding of one's judgmental processes. But it can also have no effect or even detrimental effects if decision makers hedge their choices by sticking with the status quo, picking the "middle" option, or trying to use all possible information, regardless of its diagnosticity.

Incentives

Many scholars have asked whether biases and errors would vanish if the stakes were high enough**.** Camerer and Hogarth (1999) reviewed 74 studies comparing behavior of experimental subjects who were paid zero, low, or high financial incentives based on their performance. Some tasks show performance improvement with higher incentives, such as the recall of items (Libby & Lipe, 1992), the effects of anchoring on judgment, the solving of easy problems, and clerical tasks.

People often work harder with incentives, but if the task is too complicated, incentives can hurt rather than help. Without the relevant cognitive training, participants may succumb to the "lost pilot" effect – I don't know where I am going, but I'm sure making good time (Larrick, 2004). Incentives hurt performance when extra effort is used inappropriately. In one study, Arkes, Dawes, and Christensen (1986) asked subjects to predict whether 20 students would or would not win honors given their grades. Participants were given a simple formula that gave the correct answer 70% of the time. Without incentives, subjects used the formula and got 66% correct. With incentives, subjects tried more complex rules and did slightly worse (63% correct).

Most of the studies showed that incentives had no effect on performance. Camerer and Hogarth (1999) concluded that "incentives can reduce self-presentation effects, increase attention and effort, and reduce thoughtless responding," but "no replicated study has made rationality violations disappear purely by raising incentives" (p.7).

Calibration Training

As discussed earlier, most people are overconfident, even when it is in their best interests to be well-calibrated. Outcome feedback may be unclear, delayed, or even nonexistent. Only two groups of professionals studied so far have accurate assessments of their own knowledge: meteorologists (Murphy & Winkler, 1977) and racetrack handicappers (Ceci & Liker, 1986). Their ability may stem from the fact that they face similar problems every day, make precise probability estimates, and receive quick and clear feedback.

Firms can suffer severe financial consequences from overconfidence. Some years ago, Royal Dutch/Shell noticed that their newly hired geologists were wrong much more often than their level of confidence implied, costing Shell millions of dollars with their overconfidence. For instance, junior geologists would estimate a 40% chance of finding oil in a particular location, but when ten wells were drilled, only one or two would produce oil.

The geologists needed better feedback. Shell designed a training program that used feedback to help its geologists become better calibrated. Using past cases of drilled wells, geologists were presented with many factors affecting oil deposits. For each case they had to provide their best guesses for the probability of striking oil as well as ranges for how much a successful well might yield. Then they received feedback. The training worked, and Shell geologists are now better calibrated (Russo & Schoemaker, 1992). Similar procedures could be used in medicine, the stock market, legal settings, and other domains where miscalibration could have serious consequences.

Conclusion

If our judgments are really this bad and methods of debiasing only help some of the people some of the time, how in the world do we manage to make reasonably good decisions most of the time? The psychological literature is filled with controversy about the heuristics and biases framework.

According to one type of critique, the alleged biases are not really biases at all. Errors are defined relative to normative theories, and normative theories rest on assumptions about decision makers' goals, such as maximizing expected utility or minimizing squared errors. If people care about other things, such as obeying social norms, maintaining good impressions, or sticking to their principles, their behavior is obviously not erroneous (Gigerenzer & Murray, 1987). Dsyfunctional errors in one framework may be perfectly reasonable in another (Tetlock, 2002).

Other critiques of the heuristics and biases framework focus on the generality of errors across people and situations. Biases may occur in laboratory studies, but real world competitive markets should reduce their occurrence. Furthermore, not everyone is equally vulnerable to errors. It helps to be smart. Cognitive ability is correlated with fewer biases in judgment, including the ability to overcome framing effects and the conjunction fallacy (Stanovich & West, 1998). It also helps to be knowledgeable. Statistical training reduces some biases (Nisbett, Fong, Lehman, & Cheng, 1987).

Still other critiques of the heuristics and biases approach focus on inadequate theory. Heuristics help us organize and understand effects, but what good are they if the findings can be explained by multiple heuristics? Take base rate neglect, the tendency for people who make probabilistic inferences to discount base rates in favor of individuating information. Base rate neglect was initially illustrated with the cab problem (Tversky & Kahneman, 1982) which read:

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. 85% of the cabs in the city are Green and 15% are Blue. A witness identified the cab as "Blue". The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness could correctly identify the color of cab 80% of the time and failed to identify the correct color 20% of the time. What is the probability that the cab involved in the accident was Blue rather than Green?

The Bayesian solution to this problem is 41%, although most people say the probability that a Blue cab was involved in the accident is closer to 80% (the individuating information). Kahneman and Tversky attributed this finding to the representativeness heuristic according to which probability is substituted for similarity. Yet the tendency to say 80% might also be attributed to the availability heuristic if the witness's statement was more vivid and salient than the base rate. Or the tendency to say 80% might be due to the anchoring and adjustment heuristic if people anchor on the individuating information and insufficiently adjust for base rate information.

Finally, other critiques of heuristics and biases say the pendulum has swung too far. With the focus on biases and errors, this research agenda has skewed our perception of human behavior. People have amazing powers, such as the ability to learn patterns, chunk information, store it, and retrieve it as needed. In an ingenious study, Chase and Simon (1973) presented a master chess player and a novice with a half-finished game of chess for 5 seconds. The chess board contained approximately 25 pieces. The master and the novice were then asked to reconstruct the board. The master was able to correctly locate an average of 16 pieces, while the novice located only 4 pieces correctly. These master chess players do not perceive *all* patterns better. Chase and Simon briefly presented the master and the novice with another chess board containing 25 randomly placed pieces. Both the expert and novice were able to locate only 3 or 4 pieces correctly. The chunking, storing, and retrieving that experts have mastered applies only to meaningful patterns.

Another form of pattern matching is facial recognition. Human faces are composed of identical components, and yet peoples' memory for specific faces is remarkable. In one study, Bahrick, Bahrick, and Wittlinger (1975) tested the memory of people for their high school classmates. They obtained the high school yearbooks of approximately 400 people who ranged in age from 18 to 74. Ten pictures were randomly selected from each yearbook and photocopied. Each picture was placed on a card with four other pictures randomly taken from other yearbooks. Participants were shown the cards and asked to identify the photo of their classmate. Choices were correct 90% of the time, even among participants who had graduated 35 years earlier. Those who had graduated 48 years ago were *still* correct 71% of the time. It is easy to imagine evolutionary reasons for our remarkable talent to recognize faces.

Individuals may have some well-tuned cognitive skills such as pattern matching, but what about groups? Are they capable of any amazing feats? Surowiecki (2004) describes conditions under which groups of individuals outperform the elite few. He tells a story of Francis Galton, who came across a weight-judging contest at a country fair. Individuals entered their guess of the weight of an ox, and the best guesses received prizes. Approximately 800 people entered the contest. When it was over, Galton borrowed the tickets from the organizers and plotted the estimates to see if they formed a bell curve. In the process, he calculated the mean, and to his surprise, the average guess of the ox's weight was 1,197 pounds—only one pound away from the actual weight of 1,198 pounds. Even when members of a group are not particularly well-informed, the group can reach a collectively wise decision, provided the individual opinions are independent, diverse, and decentralized (i.e., no one is at the top dictating the right answer). In addition, there must be a reasonable method for aggregating judgments and determining the group decision.

Both experts and novices are susceptible to a host of biases, and debiasing does not always help. Nonetheless, we can learn from our own mistakes and from those of others. As Sam Levenson once said, “We must learn from the mistakes of others. We can’t possibly live long enough to make them all ourselves.”

References

Arkes, H.R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*, 125-140.

Arkes, H.R., Dawes, R.M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes, 27,* 148-180.

Arkes, H.R., Saville, P.D., Wortmann, R.L., & Harkness, A.R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology*, *66*, 252-254.

Armstrong, K., Schwartz, J.S., Fitzgerald, G., Putt, M., & Ubel, P.A. (2002). Effect of framing as gain versus loss on understanding and hypothetical treatment choices: Survival and mortality curves. *Medical Decision Making, 22,* 76-83.

Asch, S.E. (1946). Forming impressions of personality. *Journal of Abnormal Psychology, 41,* 258-290.

Ashton, R. H. (1992). Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes, 52*, 292-306.

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives, 11,* 109-125.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104,* 54-75.

Baker, L.A., & Emery, R.E. (1993). When every relationship is above average: Perceptions and expectations of divorce at the time of marriage. *Law and Human Behavior, 17,* 439-450.

Barber, B. & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, *55*, 773-806.

Bukszar, E. & Connolly, T. (1988). Hindsight bias and strategic choice – Some problems in learning from experience. *Academy of Management Journal*, *31,* 628-641.

Cain, D.N., Loewenstein, G., & Moore, D.A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies, 34*, 1-25.

Camerer, C. F. & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*, 7-42.

Ceci, S.J. & Liker, J.K. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology: General, 115*, 255-266.

Chase, W.G., & Simon, H.A. (1973). The mind's eye in chess. In W.G. Chase (Ed.) *Visual information processing* (pp.215-281). New York: Academic Press.

Chapman, G.B., & Johnson, E.J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making, 7,* 223-242.

Dellarosa, D., & Bourne, L.E. (1984). Decisions and memory: Differential retrievability of consistent and contradictory evidence. *Journal of Verbal Learning and Verbal Behavior, 23*, 669-682.

Desvouges, W.H., Johnson, F., Dunford, R., Hudson, S., Wilson, K., & Boyle, K. (1993). Measuring resource damages with contingent valuation: Tests of validity and reliability. In *Contingent valuation: A critical assessment.* Amsterdam: North Holland.

Fischhoff, B. (1975). Hindsight ≠ Foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288-299.

Fischhoff, B. & Beyth, R. (1975). “I knew it would happen”: Remembered probabilities of once-future things. *Organizational Behavior and Human Decision Processes*, *13*, 1-16.

Fischhoff, B. Slovic, P., & Lichtenstein, S. (1986). Knowing with certainty: The appropriateness of extreme confidence. In H.R. Arkes & K.R. Hammond (Eds.) *Judgment and decision making: An interdisciplinary reader*. (pp. 397-416). New York: Cambridge University Press.

Frisch, D. (1993). Reasons for framing effects. *Organizational Behavior and Human Decision Processes, 54,* 399-429.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684-704.

Gigerenzer, G., & Murray, D.J. (1987). *Cognition as intuitive statistics.* Hillsdale, NJ: Erlbaum.

Gourville, J.T., & Soman, D. (1998). Payment depreciation: The behavioral effects of temporally separating payments from consumption. *Journal of Consumer Research, 25,* 160-174.

Hoch, S.J. (1985). Counterfactual reasoning and accuracy in predictive judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 719-731.

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine, 73,* 538-540.

Huber, J., Payne, J.W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9,* 90-98.

Johnson, E.J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty, 7,* 35-51.

Joyce, E.J., & Biddle, G.C. (1981). Anchoring and adjustment in probabilistic inference in auditing. *Journal of Accounting Research, 19,* 120-145.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribution substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.). *Heuristics and biases: The psychology of intuitive judgment* (pp.49-81). New York: Cambridge University Press.

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,* 203-235.

Kidd, J.B. (1970). The utilization of subjective probabilities in production planning. *Acta Psychologica, 34,* 338-347.

Knetsch, J.L. (1989). The endowment effect and evidence of nonreversible indifference curves, *American Economic Review, 79,* 1277-1284.

Kruglanski, A.W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology, 32,* 437-459.

Larrick, R. (2004). Debiasing. In D. J .Koehler and N. Harvey (Eds.) *Blackwell handbook of judgment and decision making* (pp.316-377). Malden, MA: Blackwell Publishing.

Larwood, L. & Whittaker, W. (1977). Managerial myopia: Self-serving biases in organizational planning. *Journal of Applied Psychology, 62,* 194-198.

Leary, M.R. (1982). Hindsight distortion and the 1980 presidential election. *Personality and Social Psychology Bulletin*, *8,* 257-263.

LeBoeuf, R.A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making, 16,* 77-92.

Lerner, J. S. & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125,* 255-275.

Levin, I.P., & Gaeth, G.J. (1988). Framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15,* 374-378.

Levin, I.P., Gaeth, G.J., Schreiber, J., & Lauriola, M. (2002). A new look at framing effects: distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes, 88,* 411-429.

Libby, R. & Lipe, M.G. (1992) Incentives, effort, and the cognitive processes involved in accounting-related judgments. *Journal of Accounting Research, 30,* 249-273.

Loewenstein, G., Issacharoff, S., Camerer, C., & Babcock, L. (1993). Self-serving assessments of fairness and pretrial bargaining. *Journal of Legal Studies, XXII,* 135-159.

Lusted, L.B. (1977). *A study of the efficacy of diagnostic radiologic procedures: Final report on diagnostic efficacy*. Chicago: Efficacy Study Committee of the American College of Radiology.

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.A, & Cooke, A.D.J. (1996). The role of task and context in preference measurement. *Psychological Science, 7*, 76-82.

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

Messick, D.M. & Bazerman, M.H. (1996). Ethical leadership and the psychology of decision making. *Sloan Management Review, 37,* 9-23.

Messick, D., & Sentis, K. (1983). Fairness, preference, and fairness biases. In D.M. Messick & K.S. Cooke (Eds.) *Equity Theory: Psychological and Sociological Perspectives.* New York: Praeger.

Murphy, A.H., & Winkler, R.L. (1977). Can weather forecasters formulate reliable probability forecasts of precipitation and temperature? *National Weather Digest, 2*, 2-9.

Neale, M.A., and Bazerman, M. (1992). Negotiating rationally: The power and impact of the negotiator's frame. *Academy of Management Executive, 6*, 42-51.

Nisbett, R.E., Fong, G.T., Lehman, D.R., & Cheng, P.W. (1987). Teaching reasoning. *Science, 238*, 625-631.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, *29*, 261-265.

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Commission bias and ambiguity. *Journal of Behavioral Decision Making, 3,* 263-277.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37,* 332-337.

Russo, J.E., & Schoemaker, P.J.H. (1989). *Decision traps.* New York: Simon and Schuster.

Russo, J.E., & Schoemaker, P.J.H. (1992). Managing overconfidence. *Sloan Management Review, 33,* 7-17.

Samuelson, W. & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1,* 7-59.

Siegel-Jacobs, K. & Yates, J.F. (1996). Effects of procedural and outcome accountability on judgmental quality. *Organizational Behavior and Human Decision Processes, 65,* 1-17.

Simon, H.A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research, 16,* 158-174.

Simonson, I. & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes, 51,* 416-429.

Sloman, S.A., Over, D., Slovak, L., & Stibel, J.M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91,* 296-309.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D.G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis, 24*, 1-12.

Sporer, S.L., Penrod, S.D., Read, J.D., & Cutler, B.L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315-327.

Stael von Holstein, C.-A.S. (1972). Probabilistic forecasting: An experiment related to the stock market. *Organizational Behavior and Human Performance, 8*, 139-158.

Stanovich, K.E., & West, R.F. (1998). Individual differences in framing and conjunction effects. *Thinking and Reasoning, 4,* 289-317.

Staw, B.M. (1976). Knee-deep in big muddy: Study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance, 16*, 27-44.

Staw, B.M., Barsade, S.G., & Koput, K.W. (1997) Escalation at the credit window: A longitudinal study of bank executives' recognition and write-off of problem loans. *Journal of Applied Psychology, 82,* 130-142.

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.

Suroweicki, J. (2004). *The wisdom of crowds.* New York: Doubleday.

Synodinos, N.E. (1986). Hindsight distortion: I knew it all along and I was sure about it. *Journal of Applied Social Psychology*, *16,* 107-117.

Tetlock, P.E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45*, 74-83.

Tetlock, P.E. (2002). Social-functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109,* 451-471.

Tetlock, P.E., & Boettger, R. (1989). Accountability amplifies the status quo effect when change creates victims. *Journal of Behavioral Decision Making, 7,* 1-23.

Tetlock, P.E., & Kim, J.I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology, 52,* 700-709.

Tetlock. P.E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology, 57,* 632-640.

Thaler, R. (1999). Mental accounting matters. *Journal of Behavioral Decision Making, 12,* 183-206.

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). Cambridge, England: Cambridge University Press.

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

Wagenaar, W.A., & Keren, G.B. (1986). Calibration of probability assessments by professional blackjack dealers, statistical experts, and lay people. *Organizational Behavior and Human Decision Processes, 36,* 406-416.

Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129-140.

Weldon, E., & Gargano, G.M. (1988). Cognitive loafing: The effects of accountability and shared responsibility on cognitive effort. *Personality and Social Psychology Bulletin, 8,* 226-232.

---

[1] Kahneman and Frederick note that anchoring does not fit this model since it does not involve substituting one attribute for another.