Annual Review of Psychology

Judgment and Decision Making

Baruch Fischhoff[1] and Stephen B. Broomell[2]

[1]Department of Engineering and Public Policy, and Institute for Politics and Strategy, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; email: baruch@cmu.edu

[2]Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; email: broomell@cmu.edu

Annu. Rev. Psychol. 2020. 71:331–55

First published as a Review in Advance on July 23, 2019

The *Annual Review of Psychology* is online at psych.annualreviews.org

https://doi.org/10.1146/annurev-psych-010419-050747

Keywords

judgment, decision making, preference, risk, uncertainty, history

Abstract

The science of judgment and decision making involves three interrelated forms of research: analysis of the decisions people face, description of their natural responses, and interventions meant to help them do better. After briefly introducing the field's intellectual foundations, we review recent basic research into the three core elements of decision making: judgment, or how people predict the outcomes that will follow possible choices; preference, or how people weigh those outcomes; and choice, or how people combine judgments and preferences to reach a decision. We then review research into two potential sources of behavioral heterogeneity: individual differences in decision-making competence and developmental changes across the life span. Next, we illustrate applications intended to improve individual and organizational decision making in health, public policy, intelligence analysis, and risk management. We emphasize the potential value of coupling analytical and behavioral research and having basic and applied research inform one another.

Contents

INTRODUCTION

Behavioral decision research arose as psychology's contribution to a remarkable period in which scientists and scholars from diverse disciplines collaborated in pursuing issues raised by von Neumann & Morgenstern (1944) in their landmark volume on rational choice theory. In a seminal article and a subsequent *Annual Review of Psychology* article, Edwards (1954, 1961) framed the field's fundamental commitment to studying the properties of tasks, as presented by life or by researchers, in tandem with studying individuals' responses to them. The present review emphasizes the integrative strategy of those groundbreaking works and its contribution to interventions designed to help people navigate worlds that can be unfamiliar, uncertain, unintuitive, and unfriendly.

Early descriptive research focused on highly structured tasks amenable to analysis that identified optimal, or rational, behavior. Those studies revealed important regularities, such as individuals' lack of insight into their own decision-making processes and their tendency to extract too much information from some observations and too little from others. However, the research also revealed limits to such tasks, summarized in what proved to be a terminal review of two prominent early research programs (Slovic & Lichtenstein 1971). One of those programs used multiple regression to model repeated decisions (e.g., evaluating graduate school applicants based on a common set of attributes). Although those models often had predictive value, they could not distinguish among competing psychological accounts (Jaccard 2012). The second of those research programs used stylized Bayesian tasks (e.g., assessing the probability that an urn contained 70% blue balls, rather than 70% red ones, based on a sequence of draws). Although those studies revealed much about experimental design, their tasks lacked the realism needed to engage natural processes.

Recognizing these limitations opened the door to research programs that allowed richer expressions of behavior and deeper understanding of the processes producing it, while still being grounded in task analysis. One such program is the study of judgment heuristics, informed by

Bayesian analysis of the biases that they can produce (Tversky & Kahneman 1974). A second is the study of orderly but nonrational choice processes, informed by utility theory analyses of the inconsistent preferences that they produce (Kahneman & Tversky 1979). A third is the study of cognitive capabilities, informed by modeling procedures that accommodate properties of people and tasks (Karelaia & Hogarth 2008, Lieder et al. 2018). A propelling force in these developments has been the adoption of more diverse research methods, including ones that can reveal when people view tasks very differently than researchers imagine (Medin et al. 2017).

Our review begins with basic research into the three essential elements of decision making: judgment, predicting the outcomes of choosing possible options; preference, weighing the importance of those outcomes; and choice, combining judgments and preferences to make decisions. That research asks how people, in general, behave. The next two sections describe research into two potential sources of behavioral heterogeneity: individual differences and life-span developmental changes in decision-making competence. The article then illustrates interventions designed to improve individual and organizational decision making. The concluding section describes the field's potentially productive tension between relatively stable analytical methods and ever-changing empirical results.

Utility: a latent variable, derived from the choice axioms, interpreted as representing subjective perceptions of value

Consistency: the degree to which judgments follow axioms specifying sound behavior, such as those of Bayesian inference or utility theory

Literacy test: a measure of mastery based on knowledge of test items selected to represent a domain

Mental models: beliefs about how systems behave, paralleling formal models of those systems

JUDGMENT

Sound decisions require predicting what will happen if different choices are made. The quality of those judgments can be evaluated in terms of their accuracy or their consistency. Studies of both accuracy and consistency build on analytical research formalizing these criteria. Achieving one goal need not mean achieving the other. People may have accurate beliefs about one topic but not about related ones, leading to inconsistent judgments; or they may have consistent beliefs but know very little. Both criteria continue to be central topics in behavioral decision research, or decision science, as the field is sometimes called.

Accuracy

How accurately people understand their world has been studied in several ways, each with strengths and weaknesses, as discussed below.

Knowledge (how much people know). The simplest way to evaluate how much people know is by asking them to answer factual questions. Literacy tests represent a domain (e.g., health, finance, science) with a fixed set of such questions. For example, a widely used test of science literacy asks, "True or false? The center of the Earth is very hot" (Natl. Sci. Board 2014, p. 7.23). There are many such tests. However, the selection of their items is rarely based on an analysis of what people need to know. As a result, even when scores on literacy tests predict behavior, they typically offer limited insight into how individuals acquire their knowledge or use it. Perhaps knowing the specific facts on a test helps people to make better decisions; perhaps it predicts their knowledge of other, more relevant facts; or perhaps it reflects education or test-taking ability.

A decision science approach to assessing knowledge begins by analyzing the facts needed to make specific decisions. Sometimes, people need to know just a few summary estimates, such as the risks and benefits of a medical treatment. A knowledge test for those facts might ask people to estimate the probabilities of possible outcomes (Schwartz & Woloshin 2013, Zikmund-Fisher 2019). Sometimes people need to know how things work, for example, what determines their risk of HIV/AIDS. A knowledge test for those facts might ask about how the virus can be transmitted. Such mental models of the processes determining the outcomes of decisions have been studied for domains as diverse as HIV/AIDS, climate change, contraceptives, energy consumption, pandemics, and radon (Bruine de Bruin & Bostrom 2013). Von Winterfeldt (2013) illustrates how decisions are analyzed, looking at whether to turn a baby in the breech position. Fischhoff et al. (2006) illustrate how processes are analyzed, looking at how a pandemic could unfold.

Calibration: the degree to which confidence in beliefs matches their accuracy

Scoring rules: procedures that provide incentives to report true beliefs

Sometimes, the results of mental model studies are domain specific, such as the finding that people tend to ignore herd immunity when thinking about vaccines (Downs et al. 2008) and overestimate how well they can tell whether a potential partner has a sexually transmitted infection (Downs et al. 2004). Sometimes, the results are general, such as the finding that people have difficulty predicting nonlinear processes (e.g., climate change) or how small risks mount up over time (Gonzalez & Mehlhorn 2016, Tong & Feiler 2017). Studies of mental models typically begin with open-ended interviews structured around the analysis and aimed at capturing intuitive formulations and modes of expression.

Calibration (how appropriate people's confidence in their knowledge is). Using knowledge wisely requires knowing its limits. Overconfidence can lead to making decisions without enough information and missing signs that things are going wrong. Under-confidence can lead to the opposite. Perhaps the most common way to study the appropriateness of confidence is with calibration tests. These tests ask people to indicate the probability that they have answered each question in a set correctly. People are perfectly calibrated when they are correct *x*% of the time when giving an *x*% chance of being correct. The properties of calibration tasks have been studied intensely (Budescu et al. 1997, O'Hagan et al. 2006). For example, people may believe that they have answered fewer items correctly than their item-by-item probabilities imply (e.g., they believe that they had six correct answers among ten items assigned a mean probability of 80%) (May 1991, Sniezek & Buckley 1991). As a result, global and local confidence must be assessed separately. Understanding these tasks well has allowed researchers to tailor them to studies of individual differences and training, as described below.

One recent focus of calibration research has been how to create incentives for people to reveal their true confidence and not give strategic responses (e.g., hedging, boasting). The US National Weather Service has long used scoring rules to encourage candid probability-of-precipitation forecasts (Murphy & Winkler 1974). It hopes, for example, to avoid umbrella bias, whereby forecasters overstate the probability of precipitation so that no one gets caught in the rain, even if this means many people will carry umbrellas needlessly. However, because scoring rules are so abstract, forecasters need extensive feedback to master them. As a result, scoring rules are only practical for multi-round studies like the Good Judgment Project (described below). When people cannot be trained on specific scoring rules, it may still be possible to identify the rules that they use implicitly and interpret their judgments appropriately (Merkle & Steyvers 2013).

Pooling (how much a crowd knows). In cases where individuals' knowledge is limited, and their confidence questionable, the pooled judgment of a crowd may be more accurate than that of any of its members. This topic has been the subject of intense research, including both empirical studies and formal analyses (Danileiko & Lee 2018, Davis-Stober et al. 2014, Mannes et al. 2014). In general, the research finds that crowds are more accurate when each member knows something different, meaning that their judgments are correlated with the criterion but uncorrelated with one another (or even negatively correlated). Davis-Stober et al. (2014) offer an analytical account of these conditions, including ways to evaluate the accuracy of a crowd without already knowing the correct answers for some of its predictions.

Although often called the wisdom of the crowd, such accuracy typically comes without an explanation. As a result, recipients can only guess what evidence supports a prediction, how general it is, and what it implies for their mental models of the processes producing the predicted outcomes.

Providing such explanations is an opportunity for future research, as is the related challenge of explaining the black-box predictions produced by machine learning programs that identify patterns buried in vast datasets.

Bayesian inference: a formal framework for evaluating evidence and updating beliefs, consistent with the probability theory axioms

Subjective versus frequentistic probabilities: the distinction between judgments that summarize beliefs or observations of repeated events

Consistency

The most familiar consistency standard is Bayesian inference, which provides rules for how people should evaluate evidence and update their beliefs (Edwards et al. 1963, Kyburg & Smokler 1964). It is the standard used in many well-known lines of research, such as studies that examine the conjunction fallacy (Tversky & Kahneman 1983) and base-rate neglect (Tversky & Kahneman 1974). Nonetheless, Bayesian inference has its critics. Some object to its use of beliefs, as these express subjective rather than frequentistic probabilities, which summarize the relative frequency of repeated events (e.g., coin flips, rainy days). In the long-running debate over the nature of probability, Bayesians argue that a subjective judgment is required to decide that events are identical enough to be treated as repeated. As a result, they claim that there are no objective probabilities (Edwards et al. 1963).

Other critics object to the Bayesian requirement that people allocate 100% of their subjective probability to a fixed set of hypotheses. These critics argue that people sometimes feel that their hypotheses are incomplete or unclear. In such cases, they should be able to reserve some probability for unimagined possibilities or clearer thinking (Gärdenfors & Sahlin 1988). Indeed, formal analyses have shown that some inferential tasks are so complex that having consistent beliefs may be an unreasonable aspiration (Dasgupta et al. 2017, Schum 1994).

An alternative standard of consistency, which addresses these concerns, is Dempster-Shafer inference. Rather than looking at the balance of evidence, as in Bayesian inference, Dempster-Shafer inference looks at its conclusiveness. Shafer & Tversky (1986) argue for using the consistency standard that best fits how people naturally think about a task. That advice has been followed in studies that use different standards to illuminate how people think about different kinds of evidence, often using multi-method approaches. One such study used details from a jury trial and found that people treat contradictory evidence (which says different things about the same event) differently than they treat conflicting evidence (which points in different directions) (Curley 2007). Another such study found that people have consistent beliefs about the conclusiveness of evidence, which the authors called known unknowns (Walters et al. 2017). A third study found that people use terms like "confidence" to describe uncertainty about their knowledge and terms like "likelihood" to describe their uncertainty about the world (Üklümen et al. 2016). That usage parallels Bayesians' preference for "assessing" subjective probabilities and "estimating" frequentistic ones.

These studies reflect three emerging trends, arising from the concern that people may think about tasks in fundamentally different ways than researchers imagine. One trend is using multiple tasks, hoping to triangulate on lay perspectives. The second is using open-ended tasks, letting people speak in their own terms and possibly reveal unexpected ways of thinking. The third is replacing formal constructs with approximations (e.g., known unknowns, rather than second-order probabilities), seeking the sweet spot between the questions that researchers want to ask and the questions that people can answer. For example, Walters et al. (2017) used a combination of think-aloud protocols, eliciting spontaneous expressions of uncertainty; text boxes, asking participants to write down known unknowns; and rating scales, asking for evaluations of a list of unknowns.

How people make judgments about their world has long been a central concern of decision-making research. Such research has advanced by devising tasks that allow people to reveal themselves more fully and by analyzing those tasks more thoughtfully in terms of the performance standards of accuracy and consistency. Research on preferences has progressed in much the same way, with one important difference described immediately below.

PREFERENCES

Standard gamble: a choice between a sure thing and a 50/50 chance of a better or a worse outcome, used to assess the sure thing's utility

Constructed preference: how people infer their preferences for unfamiliar choices

Decision science has no accuracy criterion for preferences. People can prefer whatever they want, an assumption that is shared by neoclassical economics. However, decision science does have a consistency standard: Preferences should follow the utility theory axioms (Edwards 1954). When that happens, people are deemed rational over the options involved. Those axioms include being able to compare any two outcomes, making trade-offs between any two outcomes, and ignoring how outcomes are described (if the end states remain the same).

Economists assume that people are rational, in this sense. They then infer what matters to them from observed behavior. Such revealed preference analyses also assume that people have stable preferences, which they reveal in all their choices; that researchers know how people perceive those choices; and, sometimes, that the choices are made in efficient markets (Becker 1976).

Psychologists are free to test these axioms, as are the behavioral economists who have followed their lead. Indeed, violations of the axioms underlie much current theory. For example, a key assumption of prospect theory (Kahneman & Tversky 1979) is that preferences depend on the reference point evoked by how outcomes are described (e.g., are raises compared with current salaries, expected raises, or other employees' raises?). Such sensitivity violates the axiom that holds that how outcomes are described should not matter, only their consequences. Another widely studied violation arises when the relative attractiveness of two options is reversed by adding a third option that is inferior to both, hence should be irrelevant. A recent review concluded that such irrelevant options have the greatest effect when people lack strong prior preferences and contextual cues are made more salient (Huber et al. 2014). The review also notes how marketers manipulate those conditions, for example, by manipulating the appearance of online reviews, using an orderly presentation to make comparisons easier or a chaotic display to make them harder.

One pitfall in preference research is that abstract axioms can lead to abstract tasks, which people have difficulty answering. Indeed, the first *Annual Review of Psychology* article on judgment and decision making lamented a study that "threw out 61 per cent of…subjects" (Edwards 1961, p. 491) for having inconsistent preferences on an abstract task that proved too confusing. Such problems have continued to plague studies that try to elicit precise preferences. For example, health-care policy analysts often pose axiom-based standard gambles, such as, "What probability of getting moderate sleep quality would be just as good as a 50/50 chance of getting the best possible or the worst possible sleep quality?" Such questions prove so hard to answer that studies routinely exclude many responses as seemingly not reflecting the respondents' true preferences (Engel et al. 2016). Similarly, cost-benefit analysts often ask people how much they are willing to pay, in dollar terms, to protect nonmarket goods (e.g., historic sites, endangered invertebrates, child welfare). These questions are so hard (or objectionable) that many people refuse to answer or give other protest responses (Meyerhoff & Liebe 2010).

From a practical perspective, these measurement failures are troubling because they undermine the credibility of the health-care policies or cost-benefit analyses that they are meant to inform. From a theoretical perspective, though, such failures can be sources of insight, showing how context affects expressed preferences. The artificiality of the tasks has encouraged a constructed preference approach (e.g., Huber et al. 2014), which assumes that people must infer their preferences for unfamiliar choices, rather than immediately knowing what they want for all possible options (as economists' stable preference assumption implies) (Lichtenstein & Slovic 2006).

Constructed preference research takes several forms. One uses experimental manipulations to compare formally equivalent tasks that evoke psychologically different processes, as in a study examining how people construct risk preferences in response to task cues (Pedroni et al. 2018). A second infers those processes from observations that capture natural variation, as in a study that observed stability in risk and time preferences but not social preferences (Chang & Schechter 2015). A third makes such inferences for experimental studies, as in a review that concluded that choice architecture field experiments that manipulate how options are presented reveal too little about participants' preference construction processes to evaluate the underlying theories (Szaszi et al. 2017).

These concerns have prompted renewed interest in process-tracing methods (Schulte-Mecklenbeck et al. 2017). These methods, which have long been part of behavioral decision research (Fischhoff 1996, Payne et al. 1993, Svenson 1979), attempt to clarify how people form preferences by asking them to think aloud or manipulate stimuli as they perform tasks. Three research trends have encouraged the adoption of such methods. One trend is the development of protocols for coding observed behavior into analytical terms (e.g., options, sources of uncertainty). These protocols allow more reliable coding, and clearer comparisons across studies, compared to the emergent codes of grounded theory, which dominate qualitative research (Bryant & Charmaz 2007). A second trend is the greater acceptance of concurrent verbal protocols in which people report how they are making decisions, thereby avoiding problems with retrospective verbal protocols in which people report how they made decisions (Ericsson & Simon 1992). A third trend is a greater willingness to accept the risks of reactive measurement, whereby researchers might influence study participants by asking them to describe their thinking, relative to the risks of misinterpreting their responses to structured tasks.

Grounded theory: qualitative research that lets theoretical categories emerge from immersion in the material

Increased methodological heterogeneity has also encouraged research into sacred (or protected) values, which people will not compromise. Such values are nonrational, because they violate the utility theory axiom that requires willingness to make trade-offs among all outcomes. However, sacred values can be central to thoughtful decisions (Baron & Spranca 1997). Mixed-method research programs have, for example, described the roles of sacred values in overcoming the psychological numbing associated with immense problems, like genocide (Slovic & Slovic 2015), and in discouraging violent extremism (Atran 2016).

Thus, with preference as with judgment, task analysis has framed descriptive research. That framing has revealed nonrational behavior worthy of theoretical accounts, such as the inconsistent preferences that prompted the development of prospect theory. It has also revealed the limits to rationality, such as the struggles with abstract tasks that prompted the constructed preference approach. Analogous patterns emerge in the study of choice tasks, wherein people combine their preferences (what they want) and their judgments (what they can get) to make decisions.

CHOICE

Birnbaum (2011) distinguishes two complementary approaches to studying how people make choices: experiments and modeling.

Experiments ask how sensitive people are to the factors that researchers manipulate. They represent a piecemeal research strategy, with each experiment estimating the effects of a few factors while holding all other factors constant. Extrapolating from any single experiment requires estimating the impact of varying each other factor. Creating a coherent account requires a suite of experiments, whose manipulations are derived from an underlying theory and supported by studies of task features (as with calibration tasks). Because recent *Annual Review of Psychology* articles have emphasized theory-driven experimental approaches (Lerner et al. 2015, Oppenheimer & Kelso 2015, Weber & Johnson 2009), we focus here on modeling.

Decision modeling uses statistical procedures such as multiple regression analysis to estimate the relative importance of the factors that describe each option in a choice set (Karelaia & Hogarth 2008). For example, the options might be graduate student apartments, with the factors being size, location, cost, and safety; or the options might be graduate students, with the factors being grade point average (GPA), graduate record examination (GRE) scores, and quality of undergraduate institution (Dawes et al. 1989). The importance of any factor depends on the set of options. For example, graduate students, who are generally sensitive to cost, might ignore it when choosing among apartments with roughly the same rent. Graduate admission committees, which normally consider GRE scores, might ignore them if they are highly correlated with GPAs. Some decision models estimate weights for synthetic factors such as loss aversion (described below), a construct central to cumulative prospect theory (CPT) (Tversky & Kahneman 1992). Given how heavily CPT has been studied, we use it to illustrate current approaches to decision modeling.

Decisions by description and by experience: the distinction between learning about choices from case-by-case observation or from summary descriptions

CPT incorporates several behavioral principles in a single model. Loss aversion is one. It reflects a tendency to be more sensitive to losses than to equal-sized gains (e.g., losing versus winning \$5). Risk tolerance, probability weighting, and choice stochasticity are other CPT principles. The CPT decision model has a parameter for each principle. Parameter values are estimated for research participants' choices among gambles described in terms of probabilities of winning and losing specified amounts. If those estimates were stable, they would give the theory predictive power. However, they have proven highly variable (Davis-Stober et al. 2016, Regenwetter & Robinson 2017).

One qualitative review concluded that the magnitude of loss aversion depended on task features such as how the outcomes are framed, how large the stakes are, and how long the experiment runs (Ert & Erev 2013). A somewhat later quantitative meta-analysis found weak overall evidence of any loss aversion (Walasek et al. 2018). However, the review's authors also lamented the poor quality of the methods and reporting in many studies, which made it unclear whether loss aversion did not exist or was lost in the noise. The decision by sampling (DbS) model estimates loss aversion (and other CPT parameters) by assuming that people make decisions by sampling their evaluations of previous options from memory and comparing them to the options in experimental choices (Stewart et al. 2006, 2015). Drawing on cognitive psychology, DbS also posits task features that can affect the sampling and comparison processes. Those features include aspects of the options (e.g., the distributions of outcomes and probabilities) and the task (e.g., the time allowed to reflect on the choice).

Sensitivity to task features means that parameter estimates may not be comparable for studies that offer different choices or present the same choices in different ways. One such task feature is whether outcomes are described, in summary form (e.g., x% chance of winning a y amount of dollars), or experienced, with people observing a set of trials before making their own choices. One proposal holds that people rely on unduly small samples when making such experience-based choices, which leads them to underweight small probabilities, contrary to the predictions of CPT, which is typically studied with description-based choices (Hertwig et al. 2004). Although initial descriptive studies appeared to support that hypothesis, a formal analysis concluded that the experience-sampling process produced different gambles than the ones described by CPT, rendering the comparison moot (Hadar & Fox 2009).

Using information theory to assess how well a set of choices can reveal decision weights, Broomell & Bhatia (2014) concluded that the stimuli commonly used to provide experience cannot, in principle, be used to estimate the CPT parameters. As a result, those stimuli could not reveal the underweighting of small probabilities, even if it were to occur. This analytical approach has allowed reanalysis of existing studies that compared decisions by description and by experience (Kellen et al. 2016) and has guided the design of tasks that could allow estimating decision weights for experience-based choices in studies that found less sensitivity to probabilities than with description-based choices (Glöckner et al. 2016).

Like most behavioral decision research, description-based choices involve one-time decisions. Experience-based choices revive the field's early interest in repeated decisions, including both sequential decisions, in which information accumulates over time, and dynamic decisions, in which choices can affect the options faced in future rounds. However, as noted in the third *Annual Review of Psychology* article on decision-making research, determining the optimal solution for repeated choices can be daunting for researchers and impossible for research participants (Rapoport & Wallsten 1972). An alternative research strategy engages people in multiple-play computer simulations and then compares their behaviors with the results of having the computer apply well-defined choice strategies. Soon after such simulations became technically possible, Brehmer (1992) and his colleagues created one for fighting forest fires, which appeared to engage its sponsors, the Swedish Armed Forces, which could see analogies with their own domain, without quibbling about technical (military) details. The price to pay for such verisimilitude is having to derive solutions experientially rather than analytically (Kahneman & Klein 2009). More recent dynamic decision-making research has linked tasks to theories of cognitive processes (Gonzalez & Mehlhorn 2016, Mohan et al. 2017).

Decision-making competence (DMC): mastery of judgment, preference, and choice skills, as measured by experimental tasks

Thus, with choice, as with judgment and preference, the commitment to characterizing tasks in analytical terms has allowed researchers to pool results across diverse tasks, revealing both general trends and variation. The next two sections consider research addressing two possible sources of variation: individual differences and life-span changes in decision-making competence. These studies, too, reflect the increased heterogeneity of the field's methods, tasks, and perspectives.

INDIVIDUAL DIFFERENCES

Individual differences played little role in early behavioral decision research. One reason was that researchers focused on how people, in general, behave. That focus encourages research that varies tasks across studies rather than standardizing them, as required for individual-difference measures. A second reason was that the tasks were not understood well enough to take advantage of the precision of decision science constructs, compared to the bewildering richness of constructs for personality (Ashton et al. 2004) and cognitive style (Pashler et al. 2009). A third reason was that early studies found so little evidence of individual differences in risk-taking propensity (Slovic 1964) and cognitive style that Huber (1983) recommended abandoning the search, absent a breakthrough in theory or method (and then followed his own advice).

One such breakthrough arose from recognizing that people who take risks in one domain (e.g., health, sports, research) need not take them in others (e.g., investment, child care). That insight underlies the Domain-Specific Risk-Taking (DOSPERT) scale, which can be adapted to specific domains in ways that facilitate comparisons across them (Weber et al. 2002). The Medical Maximizer-Minimizer Scale (MMS) focuses on a single domain, asking whether people describe themselves as trying to find the best possible option or just an adequate one when making medical decisions (Scherer et al. 2016). Jackson et al. (2017) offer a battery of measures assessing both decision-making style and performance.

Our own research, developing individual-difference measures of decision-making competence (DMC), illustrates such studies. Our measures used tasks selected from experimental studies of judgment, preference, and choice. Those tasks used both accuracy and consistency performance standards and differed enough to reduce shared-method variance (Podsakoff et al. 2012). A youth version (Y-DMC) was administered at the age-18 assessment to participants in a longitudinal study of the Center for Drug and Alcohol Research (CEDAR), which followed them from age 10 to age 30 (Tarter & Vanyukov 2001). Scores on the main Y-DMC factor correlated with CEDAR measures in ways that affirmed the tasks' external validity—and, by implication, that of the research literature from which they were drawn (Parker & Fischhoff 2005).

Y-DMC scores were higher for CEDAR participants who were fortunate enough to have grown up in conditions that might model and reinforce good decision making, including higher socioeconomic status, greater social support, more positive peer environments, and lower risk status (defined as not having a father with a substance abuse problem). Y-DMC scores were lower for CEDAR participants who behaved in ways that suggest poor decision making. Those behaviors included antisocial disorders, delinquency, marijuana use, and having multiple sexual partners. Y-DMC scores were also higher for participants with higher scores on tests of fixed and fluid intelligence. However, the general patterns remained in semi-partial correlations controlling for those scores. An adult version of the measure (A-DMC) showed similar patterns (Bruine de Bruin et al. 2007). When administered to CEDAR participants at their age-30 assessment, scores on A-DMC and Y-DMC (from age 18) correlated 0.50, suggesting stable individual differences (Parker et al. 2018).

In these studies, neighborhood disadvantage (at age 10) was the strongest predictor of both Y-DMC and A-DMC scores, both with and without controlling for the intelligence scores. That result is consistent with the diverse evidence that Mullainathan & Shafir (2013) assembled in arguing for the pervasive negative effects of resource constraints on decision making. Understanding the role of such social factors in decision making is an important topic for future research. For example, how do perceptions of opportunity and discrimination affect how people acquire and apply their decision-making skills? How do their decisions reflect their perceived ability to recover from the misfortune that sometimes awaits even the best decisions and decisions that require making the best of a bad situation (Hall et al. 2014)?

CEDAR provided an unusual opportunity to track changes over time. The next section describes cross-sectional research, examining the developmental course of decision making by comparing individuals in different age cohorts.

LIFE SPAN

The correlation between DMC scores at ages 18 and 30 showed consistency in relative performance. However, the two tests (Y-DMC and A-DMC) were sufficiently different that we did not compare absolute scores and therefore did not ask how much better (or worse) CEDAR participants had become as decision makers over that period. However, an increasing number of studies have administered the same tasks to people of different ages and then compared their performance. Researchers have focused in particular on adolescents, hoping to help them survive vulnerable years, and on the elderly, hoping to help them live out their lives with dignity (Hess et al. 2015, Reyna et al. 2012).

One widely cited claim about adolescents echoes folk wisdom in holding that they have an irrational sense of personal invulnerability (Elkind 1967). In examining such claims, behavioral decision research begins by analyzing teens' decisions. That analysis can clarify when and why teens and adults see decisions differently and make different choices, given their different beliefs and preferences. It can also clarify how teens' decision making interacts with other aspects of their lives. For example, impulsiveness, sometimes linked to the teen brain, might undermine teens' decision making, if it leads them to act against their own best judgment. However, poor decision making might also invite impulsiveness, if it leads teens to drift indecisively from situations where deliberation is possible to situations where emotions dominate.

In an example of a study adopting this perspective, Goldberg et al. (2009) had teens judge the risks and benefits of trying marijuana. A priori, teens who decide to try marijuana might be impulsive or have an exaggerated sense of personal invulnerability. However, the best predictor of trying marijuana proved to be whether teens believed that marijuana would prove so good that they could not stop using it. Teens who did not realize that possibility would have a failure of affective forecasting. A health message for reducing that risk might, paradoxically, emphasize how unimaginably good marijuana can be for some people, who cannot know until they try.

In addition to offering analytical tools, theoretical perspectives, and measurement methods that complement other approaches to studying teens, behavioral decision research also imposes a discipline: Understanding any decision begins by analyzing how fully informed individuals would view it in terms relevant to their values (Fischhoff 1996, 2008). That discipline can reveal issues that might otherwise be missed. For example, by identifying what they called the "risk in the benefit" of marijuana, Goldberg et al. (2009) raised the question of how to convey the addictive potential of marijuana to people who have never tried it. Would it help to point to other behaviors that some people find too good to stop (e.g., drinking, smoking, exercising, eating donuts)? What do people infer when others claim that they could stop but never do? What are the risks and benefits of teens' heightened sensitivity to peers' feelings and responses?

The discipline of analyzing decisions can also reduce the risk of a rush to judgment when assessing the competence of individuals whose choices appear suboptimal. The stakes riding on teens' perceived ability to make sound choices can be high (Blakemore 2018, Casey 2013, Salekin 2015). Justice Antonin Scalia criticized the American Psychological Association (APA) for highlighting teens' competence in a case regarding reproductive rights and teens' incompetence in a case regarding adjudication as adults for violent offenses [*Roper v. Simmons* (2005), dissenting opinion].

In principle, both APA claims could be valid (Steinberg et al. 2009). The teens and the decisions in the two cases are very different. However, evaluating those differences requires detailed analysis, considering teens' options, goals, beliefs, and constraints before making general claims about their DMC and affective control. That analysis would ask, for example, how social coercion affects teens' options, how trustworthy their information sources are (and appear to be), and what safety net backstops the experimentation essential to their development. A decision science perspective could also help clarify the self-regulatory processes studied by developmental psychologists (Blakemore 2018, Casey 2013).

The promise of such collaboration can be seen in an application of fuzzy-trace theory (Reyna 2012), which asks how people extract the gist of a decision, to a sexual behavior program (Reyna & Mills 2014). Teens' understanding improved when the program added modules giving the gist of cognitively difficult issues [e.g., "Even low risks add up to 100% if you keep doing it" (Reyna & Mills 2014, p. 1633)]. Another example is applying decision science principles to the forensic evaluation and treatment of juveniles (Salekin 2015). In these examples, as in the interventions described below, decision science addresses only one element of a complex setting. If basic research results fail to replicate, it could mean that they are not true or that they are overwhelmed by factors held constant in basic research, like orchids that wilt outside restricted conditions.

Studies of aging have also surged in recent years, prompted by concern for an aging population and aided by the relative ease of studying older people compared to teens. Whereas studies of teens have focused on how they acquire decision-making skills, studies of aging have focused on how people lose them (Levy et al. 2018). Whereas teens are often viewed as failing or flailing in a supportive world, aging adults are often seen as struggling in a hostile one (Ross et al. 2014).

Here, too, shared concerns have prompted unusual collaborations among fields as varied as neuroscience, learning, memory, intelligence, emotion, health behavior, and dyadic relationships (Hess et al. 2015). Here, too, discerning the roles of multiple processes can be challenging. One integrative study used structural equation modeling to assess how performance on decision-making tasks was related to changes in sensory functioning, processing speed, and education. It found that age-related decline in working memory was a strong predictor of performance decrements, even after controlling for other factors (Del Missier et al. 2015).

Age groups are heterogeneous, meaning that caution is needed when generalizing about them. With that proviso, current results might support some guarded conclusions: By the mid-teen years (15–16 years of age), adolescents appear to have acquired the (imperfect) cognitive decision-making skills of adults. They possess knowledge of varying quality, depending what they have experienced and been taught, and whose word they trust. They have less control over their emotions and social environment, potentially compromising the balance of reason and passion in their choices. They have greater need and desire for experience and experimentation (Reyna et al. 2012).

As people age, they appear to retain their basic decision-making skills, barring health-related impairment. However, their proficiency in applying those skills may decline for tasks requiring complex mental operations. On the other hand, for familiar decisions, they may have learned what to choose and how to live with the outcomes. They may, however, be just as vulnerable for unfamiliar ones (Hess et al. 2015, Ross et al. 2014).

These collaborations between decision science and developmental psychology appear mutually beneficial. The former offers analytical methods for characterizing decisions and theoretical perspectives for interpreting behavior. The latter offers understanding of decisions' social, affective, and physiological context. Advances in neuroimaging and comparative (interspecies) psychology have spurred research using decision-making tasks suited to those research settings (Blakemore 2018, Casey 2013). Deeper involvement of decision scientists might aid in interpreting results and designing tasks.

One topic for future research is how the decisions that people face vary across the life span. That research could ask when teens' apparent failings reflect not less decision-making competence but more difficult choices, as they learn to deal with school, careers, relationships, sexuality, avocations, drugs, alcohol, and more. In that light, the fairest intergenerational comparisons might be with major new choices. For older people, those choices might include retirement, serious illness, downsizing, and loss. A second promising area is research on children and infants, which lost two creative researchers in their prime, Janet Jacobs (Jacobs & Klaczynski 2005) and Vittorio Girotto (Girotto & Gonzalez 2008). A third is applying decision science to the elusive concept of wisdom as expressed at different ages (Grossmann 2017).

APPLICATIONS

Improving Decisions and Decision Making

Decision science interventions seek to empower people to make sound, independent choices and to provide needed protections when that proves impossible. Its interventions can be evaluated in two ways. One is seeing whether they lead to people making better choices. The second is seeing whether they lead to people having better decision-making processes, from which better choices should follow.

The first strategy also underlies libertarian paternalist interventions, which manipulate individuals' choice architecture to induce better choices, defined as those that would be made by fully informed, rational individuals (Thaler & Sunstein 2008). Such interventions would, for example, make organ donation the default option only for people whose survivors will accept that choice without having had a family consultation. They would invoke social norms to encourage health behaviors (e.g., diet, vaccinations) only for people who have the resources to adopt them and a safety net should things go wrong. They would direct retirement savings to the stock market only for people whose expected financial returns outweigh the expected psychological cost from experiencing market corrections and the economic risk from being in the stock market when the funds are needed.

Analyzing each target individual's decision is, however, too demanding for most interventions. Medical decision-making researchers pursue a more modest but still ambitious goal: identifying the best choices for modal patients (Schwartz & Bergus 2008). They use standard gambles to elicit health state preferences and combine them with medical knowledge to identify the choices that fully informed, rational patients would make. They address patient heterogeneity with sensitivity analyses, repeating the calculations with values drawn from the distributions of patient conditions and preferences (Basu & Meltzer 2007). They may also create decision aids, letting patients explore the decision space themselves (Ott. Hosp. 2019).

In order to measure health states better, the National Institutes of Health has created an inventory of psychometrically validated self-report scales, available online at no cost, with adaptive testing for efficient administration (Cella et al. 2007). That initiative, called PROMIS®, was prompted by a proliferation of outcome measures of widely varying quality, which had reduced comparability across studies (e.g., different ways to elicit self-reported pain or cognitive functioning). More recently, PROMIS has applied decision science methods to estimate utilities for seven of its domains (e.g., sleep quality, social functioning) for use in health-care policy analyses (Dewitt et al. 2018).

The alternative to promoting better choices is promoting better decision-making processes, as defined by performance on judgment, preference, and choice tasks. Such interventions have been tried ever since researchers realized that people are imperfect decision makers (Slovic et al. 1977). One natural strategy is warning people about biases. Unfortunately, such warnings appear to have limited value (Milkman et al. 2009). People may lack the cognitive structures or capacity needed to act on them; or they may neglect warnings in situations that evoke intuitive, rather than reflective, decision making. They may also consider themselves immune to bias, once they have learned about the error from observing others' behavior (Kahneman 2011, Kahneman & Klein 2009).

However, it has long been known that people can master some skills when provided the conditions needed for learning: prompt and unambiguous feedback, proper incentives, and instruction in unintuitive processes (e.g., how risks mount up over time). Individuals who have such conditions, such as weather forecasters (Murphy & Winkler 1974) and financial auditors (Tomassini et al. 1982), have sometimes been found to produce reasonably well-calibrated confidence assessments. The next two sections describe interventions designed to create those conditions for people who do not have them.

Confidence Assessment: The Good Judgment Project

A common finding in calibration studies is that confidence and knowledge are positively, but imperfectly, correlated, such that people tend to be overconfident with hard tasks and underconfident with easy ones. The behavioral and statistical properties of that pattern have been vigorously studied and debated (Budescu et al. 1997, O'Hagan et al. 2006). There is little controversy, though, about the poor conditions for learning that everyday experience provides. Judgments are not explicit. Feedback is delayed and scattered. Bravado may be rewarded, rather than candor. It is especially hard to accumulate the experience needed to calibrate very strong (or weak) confidence, which entails estimating small probabilities of being wrong (or right) (Wickelgren 1977).

Building on an early study (Lichtenstein & Fischhoff 1980) in which calibration improved with concentrated feedback (200 judgments per round, personal discussion of results) and generalized beyond training tasks, the Good Judgment Project created a landmark training effort (Atanasov et al. 2017, Moore et al. 2017). It recruited thousands of individuals, many with substantive expertise, to provide probabilistic forecasts for hundreds of geopolitical events and then receive structured feedback. Rather than replicating any single laboratory study, the investigators drew on any theory, method, or result that they thought might be useful. For example, they provided feedback with a scoring rule that distinguished three aspects of performance: how much people know (knowledge), how well they can distinguish levels of confidence (resolution), and how well they can assign numerical values to those levels (calibration). The study advised participants to use models in order to reduce their cognitive load and improve their reliability. It defined events precisely enough that their occurrence or nonoccurrence could be observed, as required for meaningful feedback. It also took advantage of its large sample to compare variants on its basic intervention.

Signal detection theory (SDT): a method that simultaneously estimates individuals' discrimination ability and decision rules

The Good Judgment Project found that (*a*) a brief, intense dose of training, coupled with scoring-rule feedback, produced sustained improvements; (*b*) remote interaction with other participants helped somewhat; (*c*) individual differences were stable enough to reveal superforecasters; and (*d*) people who joined, and stayed, in the study were better calibrated than participants in most previous studies (Atanasov et al. 2017, Moore et al. 2017). Given the central role of expert judgment in policy analyses (see below), these results have important practical implications (see also Dhami et al. 2015, Morgan 2017, O'Hagan et al. 2006).

Diagnostic Decisions: Night Shift

Another sustained training effort applied decision science to address a costly failure of expert judgment (Mohan et al. 2012, 2017): Despite continuing efforts by the American College of Surgeons (ACS) and others, 60% of severely injured patients who present at local hospital emergency departments (EDs) are not transferred to major medical facilities that can provide needed care. An archetypal case is an older person who has fallen, with no obvious injuries but suspected intracranial bleeding as cause or effect of the fall. Rather than transfer the patient, the local hospital orders a computerized axial tomography (CAT) scan, even though a confirmatory result will arrive after the golden period for transfer has passed.

In signal detection theory (SDT) terms (Lynn & Barrett 2014), flawed transfer decisions could reflect poor discrimination ability or poor decision rules. ED physicians might have poor discrimination ability because diagnostically difficult (as opposed to medically difficult) cases are rare and because they receive limited feedback on what happens after patients leave the ED. ED physicians might have poor decision rules because they are under financial pressure to keep patients or because they want to demonstrate their skill, not realizing that their expertise cannot compensate for their hospital's limited ability to provide aftercare. A study asking ED physicians to evaluate detailed (anonymized) patient records found that some had good discrimination but poor decision rules, whereas others had good decision rules but poor discrimination (Mohan et al. 2012).

The heterogeneity in physicians' performance means that any general intervention would have to serve physicians with diverse discrimination abilities and decision rules. To that end, Mohan et al. (2017) created two serious games designed to improve physicians' use of the representativeness heuristic (Tversky & Kahneman 1974) when assessing case severity, a criterion capturing both discrimination ability and decision rules. Presented online, both interventions sought to make atypical severe cases (e.g., an older person falling) seem more representative of the underlying pathology. Both provided feedback missing from physicians' normal experience. They differed in the learning theories that guided their design (narrative engagement versus analogical reasoning). Both improved performance on a third simulation, which was administered both immediately after training and six months later, compared to equal doses of traditional (ACS) training. Whether that improvement extends to clinical practice is an open question, and one that may be hard to answer, given the many factors affecting actual performance. In another domain, Canfield et al. (2017) found it impossible to evaluate anti-phishing training externally, despite having detailed records from spyware installed on computers (with their users' permission). They found that users' vulnerability to malware depended on factors unrelated to users' vigilance, such as their choice of computer, browser, Internet service provider, or automatic updates.

Both the Good Judgment Project and the serious games for ED physician interventions reflect the same basic learning principles: People acquire skills best when they receive good feedback about their performance, direct instruction about unintuitive patterns, opportunities to practice, and appropriate incentives. The success of both interventions required attention to myriad details, such as how well the Good Judgment Project communicated its scoring rules and how well the serious games implemented representativeness. As a result, both required collaboration with practitioners who could provide substantive knowledge (e.g., about world events, trauma), recruit expert participants, design engaging interfaces, and collect secure data. Had the interventions failed, the theory might have been flawed or its implementation might have been undermined by not getting one of those components right. The concluding section of this review discusses the conditions that foster such collaboration.

COST, RISK, AND BENEFIT ANALYSES

US Presidential Executive Order 12291 requires cost-benefit analyses for all federal policies with expected economic impacts over $100 million. Regulations in many countries require quantitative risk analyses (e.g., for policies affecting air and water pollution). Private sector organizations often commission quantitative analyses for internal or external consumption. Decision science has played three interrelated roles in such analyses: (*a*) improving the expert judgments that shape them, (*b*) translating human behavior into analytic terms, and (*c*) communicating between organizations and their stakeholders (e.g., consumers, regulators, investors, voters). The following sections illustrate each role, focusing on analyses intended to inform public policies.

Judgment in Analysis

The combination of analytical and empirical approaches allows decision science to address the two kinds of subjectivity found in any analysis: expert judgments in assessing its inputs and value judgments in setting its terms (Fischhoff & Kadvany 2011). Both contributions have benefited from early collaboration between psychologists and management scientists in creating decision analysis (Raiffa 1968, von Winterfeldt & Edwards 1986), which elicits decision makers' beliefs and preferences as inputs to calculating the expected utility of choice options. Medical decision aids (Ott. Hosp. 2019) are a form of decision analysis.

These collaborations positioned decision scientists to be active players when risk analysis emerged as a field in the 1960s and 1970s (Fischhoff 2015). They were among the founders of the Society for Risk Analysis in 1981. The US Nuclear Regulatory Commission turned to them (among others) when its technical analyses did not convince the public that nuclear power had acceptable risks (Fischhoff et al. 1981). Decision scientists were also early contributors to analyzing and communicating the risks of climate change (Chen et al. 1983).

Risk analyses decompose complex processes (e.g., nuclear power plants, terrorist attacks, sea-level rise) into more knowable parts. When data are available (e.g., valve failure rates), they can be used as model inputs. When they are not, expert elicitation can be used to provide disciplined judgments. The Good Judgment Project elicited judgments for discrete events from many individuals; more intensive methods include interactive computer programs and day-long interviews that guide experts in reflecting on the internal consistency of their judgments (Morgan 2017, O'Hagan et al. 2006). These methods assume that experts are like everyone else once they run out of evidence and must rely on judgment. Testing that assumption is an important topic for future research.

Decision science has two roles in studying the value judgments found in all analyses. The first is identifying them. For example, defining risk (or benefit) is inherently value laden. A risk analysis (e.g., for a hazardous waste facility) could consider just mortality or also morbidity. Mortality estimates could consider just the expected number of fatalities or also the expected number of life-years lost with those deaths (a measure that gives greater weight to deaths of younger individuals). An analysis could consider (or ignore) whether risks are borne by people who do not benefit from a policy; it could consider (or ignore) when deaths might occur and whether to discount future lives; it could focus on identifiable lives or statistical ones (Slovic & Slovic 2015); and so on. Decision science has helped make these issues part of public discourse and analytical practice (Fischhoff 2015, Morgan 2017).

Decision science's second role is resolving those value issues. For quantitative outcomes, it has contributed preference elicitation procedures (see above). For more qualitative aspects of risk, psychometric studies have identified dimensions of concern, with the most common being dread and uncertainty (Glassman-Fox & Weber 2016). These dimensions are often treated as irrational; however, they can be legitimate bases for public policy. Policy makers might care about the psychological (and physiological) consequences of feelings of dread (Slovic 2001). They might wonder whether lay observers' sense of uncertainty suggests problems that analysts do not see or acknowledge (Broomell & Kane 2017).

Human Behavior in Analysis

Human behavior affects the costs, risks, and benefits of many policies. However, even when analysts recognize its relevance and accept behavioral science as a source of evidence (which not all analysts do), they still need its results in analysis-friendly terms. SDT is one way to satisfy that need (Lynn & Barrett 2014). As noted above, SDT characterizes performance in terms of how well individuals can detect signals and the decision rules they use when translating those perceptions into observable behavior. For analysts, SDT provides quantitative estimates (e.g., false-negative rates) to use in their calculations, including perhaps how those estimates vary across individuals and are affected by interventions (e.g., Mohan et al. 2012, 2017).

Originally applied to vigilance tasks (e.g., "Is that radar blip a hostile aircraft?"), SDT can be used for decisions in many noisy environments. Thus, SDT estimates for triage transfer decisions could inform policies for allocating resources and reimbursing providers (Mohan et al. 2012). Swets et al. (2000) provide examples as diverse as mammography, HIV screening, and metal fatigue detection. SDT estimates could also be used as feedback (e.g., "Here's what you are missing," or "Don't be so cautious") and inputs to system design (e.g., "We need second opinions"). Canfield & Fischhoff (2018) show how SDT estimates of computer users' susceptibility to phishing attacks could inform cybersecurity risk analyses (e.g., spear-phishing for a weak link, as when John Podesta's email was hacked to gain access to the Democratic National Committee's files during the 2016 election campaign).

One application used SDT to estimate undergraduate men's discrimination ability and decision rules when assessing young women's sexual intent (Farris et al. 2008). Those estimates could inform campus policies, by showing the extent to which young men miss signals or choose to ignore them. So, too, could the finding that alcohol consumption had different effects for interpreting behavior and clothing (Farris et al. 2010). Decision science has also been used to extract the policy implications of research into the effectiveness of self-defense measures (Fischhoff 1992).

Communication

Successful policies require two-way communication, both within organizations and with those affected by their policies. Decision science has played an active role in studying and improving the content, structure, and process of such communications. Three examples will illustrate the opportunities to affect and learn from these processes.

Intelligence analysis. The Good Judgment Project's sponsorship by the Intelligence Advanced Research Project Agency (IARPA) is one example of the sporadic contacts between the intelligence community and decision science. The opportunities could be seen in a classic essay on the limits to verbal quantifiers (e.g., likely) written by Kent (1964), a founder of US intelligence analysis. One key juncture in connecting research and practice was a project examining the analytical processes contributing to Israel's vulnerability to the surprise attack in the 1973 war, conducted for Israel's Ministry of Foreign Affairs (Lanir & Kahneman 2006). The project led to a collaboration with Heuer (1999), a veteran CIA staffer who introduced decision science to the agency's training, procedures, and software. That connection facilitated sponsorship, by the director of the National Intelligence's Office of Analytical Integrity and Standards, of a National Academy of Sciences consensus report on applying decision science to intelligence analysis (Natl. Res. Counc. 2011). That report supported IARPA's behavioral initiative. NATO has a working group on communicating uncertainty in analysis (Ho et al. 2015), and the US Navy now has a chief decision scientist (Lerner 2019).

Drug regulation. When evaluating new drugs, the US Food and Drug Administration (FDA) needs transparent communication, both within the agency and with its external stakeholders (patients, providers, producers, advocates). To meet that need, FDA (2018) restructured its decision-making processes for evaluating pharmaceuticals and biologics around a benefit-risk framework grounded in decision science principles. Those principles include distinguishing between scientific and value judgments, encouraging the expression of uncertainty, and accommodating diverse forms of evidence. FDA could apply decision science because it had staff who knew the science and could translate it into agency terms. Those staff members were also instrumental in creating FDA's statutory Risk Communication Advisory Committee and its Strategic Plan for Risk Communication (Fischhoff 2017). They could not, however, overcome the regulatory inertia that stalled the adoption of a drug fact box based on decision science principles (Schwartz & Woloshin 2013).

Climate change. Although engaged with climate change since the late Carter administration (Chen et al. 1983), social, behavioral, and decision scientists had little role in communicating their results until the late George W. Bush administration—when the limits to letting the science speak for itself became painfully evident. Signs of change include *Nature Climate Change* (**https://www.nature.com/nclimate/**) as the first *Nature* publication to invite behavioral research; three National Academy of Sciences colloquia on the science of science communication, initiated by Ralph Cicerone, the Academy's president and a distinguished climate scientist; a communication guide commissioned by the Intergovernmental Panel on Climate Change (Corner et al. 2018); the prominence of behavioral research at Yale Climate Connections, Climate Central, Climate Advocacy Lab, and other initiatives (e.g., van der Linden et al. 2018); and growing climate- and energy-related research (e.g., Wong-Parodi et al. 2016).

Psychology has long played a role in implementing policies aimed at helping people to eat better, save more, stop smoking, or get along with one another. These decision science applications reflect growing roles in setting policies. Those roles include communicating public concerns to policy makers and policies to the public, constraining policies with realistic assessments of human behavior, and structuring policy-making processes. By speaking the language of policy analysis, decision science has been able to translate other psychological research into policy- and analysis-relevant terms.

CONCLUSION

After briefly describing the intellectual roots of behavioral decision research (or decision science), we have reviewed research into the essential elements of all decisions (judgment, preference, and choice), differences across individuals and the life span, and practical and policy applications. One emerging theme, most obvious in the final sections of this review, is that science and society make progress together through two bridging activities, which Baddeley (1979) called applied basic psychology (seeing how theories fare in real-world settings) and basic applied psychology (domesticating phenomena observed in those settings for basic research). A second emerging theme is that the field has increased the heterogeneity of its tasks, methods, theories, and participants, partly due to these engagements.

A distinctive feature of decision science is analyzing tasks before attempting to describe how people approach them or designing interventions. The benefits of analysis include characterizing diverse tasks in common terms, thereby allowing general patterns to emerge; having clear standards for evaluating performance (and claims of bias); and being able to communicate with people from other analytically oriented fields, such as natural scientists wary of social science and policy analysts unsure how to use behavioral evidence.

One possible limit to this strategy is creating tasks that are analytically sound but cognitively intractable (e.g., standard gambles with unfamiliar health states). However, when researchers are alert to that possibility, such failures can be productive theoretically, by prompting attempts to explain anomalous behavior, and methodologically, by prompting use of research methods that are better suited to discerning fundamental differences in how people construe tasks.

A second possible limit is placing undue emphasis on anomalies. In psychology, as in other sciences, problems can be a source of insight, as when they constrain the set of heuristics to ones that could produce a pattern of biases. However, the focus on problems can create a bias meme, whereby people are seen as the sum of their failings and their capabilities are obscured.

A third possible limit is excluding researchers who are less comfortable with analysis. Fortunately, there are many efforts to reduce barriers to entry by explaining analytical concepts in ways that emphasize conceptual, rather than technical, mastery. Lynn & Barrett (2014) offer such an introduction to SDT, as do Fischhoff & Beyth-Marom (1983) for Bayesian inference, Ert (2018) for the likelihood principle, and Kaplan (2011) for operations research. The Open Science Framework (**https://osf.io/**) and Cochrane Collaboration (**https://www.cochrane.org/**) provide accessible tutorials on many topics.

The decision science strategy, integrating analytical and behavioral research, has brought psychologists into domains that include climate change, intelligence analysis, risk management, and health-care policy. That engagement has repaid some of psychology's debt to the society that has supported it, while enriching its science with new problems, evidence, and collaborators. It has often required psychologists to play three, sometimes unusual, roles. The first role is representing all of psychology, and not just their own specialty or theory, in settings where they may be the only psychologist (or even the only scientist) present. The second is serving as translators for colleagues who find applied settings unfamiliar and perhaps even uncomfortable. The third is creating sustained relations with decision makers, in order to learn their concerns, earn their trust, benefit from their expertise, and be better able to help them. When we meet these conditions, it can be good for science and society.

SUMMARY POINTS

1. Decision science provides unique opportunities for integrating analyses of decisions and empirical studies of decision makers.
2. Decision science has gradually increased the heterogeneity of the people and tasks it studies, as well as the diversity of its methods.
3. Sustained relationships with practitioners have brought decision science into applied arenas and applied concerns into the research.
4. The study of task properties has allowed identifying tasks suited to specific uses (e.g., training, measuring individual differences, eliciting expert judgments, choice modeling).
5. Theoretically and methodologically informed interventions can improve individual and organizational decision making.
6. Analyzing tasks allows comparing individuals' abilities to the challenges facing them and protects against unsupported generalizations about their competence.

FUTURE ISSUES

1. Interest in constructed preferences will grow, prompting an increasing use of potentially reactive methods, such as think-aloud protocols.
2. Analyses of decision-making tasks will continue to improve understanding of their demands, the conditions for attributions of bias, and the opportunities for interventions.
3. Research will further disentangle the effects of experimental design (e.g., stimulus selection) on research findings.
4. Collaboration with psychologists in other fields will increase understanding of decision making over the life span.
5. Demand will increase for the application of decision science to strategic political and institutional decisions, as well as its use to inform repeated decisions.
6. Decision science will play an increasing role in helping people to explain the predictions produced by crowds, machine learning, and artificial intelligence.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Wändi Bruine de Bruin, Barry Dewitt, Deepika Mohan, and Gabrielle Wong-Parodi for comments on earlier drafts, and the Swedish Foundation for Humanities and Social Sciences and the Carnegie Corporation of New York for financial support. The views expressed are those of the authors.

LITERATURE CITED

Ashton MC, Lee K, Goldberg LR. 2004. A hierarchical analysis of 1,710 English personality-descriptive adjectives. *J. Pers. Soc. Psychol.* 87:707–21

Atanasov P, Rescober P, Stone E, Swift SA, Servan-Schreiber E, et al. 2017. Distilling the wisdom of crowds: prediction markets versus prediction polls. *Manag. Sci.* 63(3):691–706

Atran S. 2016. The devoted actor: unconditional commitment and intractable conflict across cultures. *Curr. Anthropol.* 57(Suppl. 13):S192–203

Baddeley AD. 1979. Applied cognitive and cognitive applied research. In *Perspectives on Memory Research*, ed. LG Nilsson, pp. 367–88. Hillsdale, NJ: Erlbaum

Baron J, Spranca M. 1997. Protected values. *Organ. Behav. Human Decis. Process.* 70:1–16

Basu A, Meltzer D. 2007. Value of information on preference heterogeneity and individualized care. *Med. Decis. Mak.* 27:112–27

Becker GS. 1976. *The Economic Approach to Human Behavior*. Chicago: Univ. Chicago Press

Birnbaum M. 2011. Testing theories of risky decision making via critical tests. *Front. Psychol.* 2:315

Blakemore S-J. 2018. *Inventing Ourselves: The Secret Life of the Teenage Brain*. New York: Doubleday

Brehmer B. 1992. Dynamic decision making: human control of complex systems. *Acta Psychol.* 81:211–41

Broomell SB, Bhatia S. 2014. Parameter recovery for decision modeling using choice data. *Decision* 1(4):252–74

Broomell SB, Kane PB. 2017. Public perception and communication of scientific uncertainty. *J. Exp. Psychol. Gen.* 146(2):286–301

Bruine de Bruin W, Bostrom A. 2013. Assessing what to address in science communication. *PNAS* 110(Suppl. 3):14062–68

Bruine de Bruin W, Parker A, Fischhoff B. 2007. Individual differences in adult decision-making competence (A-DMC). *J. Pers. Soc. Psychol.* 92:938–56

Reports psychometric validation of a measure assessing internal and external validity of behavioral decision research tasks.

Bryant A, Charmaz K, eds. 2007. *The SAGE Handbook of Grounded Theory*. London: Sage

Budescu DV, Erev I, Wallsten TS, Yates JF, eds. 1997. Stochastic and cognitive models of confidence. *J. Behav. Decis. Mak.* 10(3)

Review of approaches to evaluating performance on probability assessment tasks.

Canfield C, Davis AL, Fischhoff B, Forget A, Pearman S, Thomas J. 2017. *Replication: challenges in using data logs to validate phishing detection ability metrics*. Paper presented at the 13th Symposium on Usable Privacy and Security (SOUPS 2017), Santa Clara, CA, July 12–14. **https://www.usenix.org/system/files/conference/soups2017/soups2017-canfield.pdf**

Canfield C, Fischhoff B. 2018. Setting priorities for behavioral interventions: an application to reducing phishing risks. *Risk Anal.* 38(4):826–38

Casey BJ, ed. 2013. The teen brain. *Curr. Dir. Psychol. Sci.* 22(2)

Cella D, Yount S, Rothrock N, Gershon R, Cook K, et al. 2007. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap Cooperative Group during its first two years. *Med. Care* 45(5):3–11

Chang Y, Schechter L. 2015. Stability of experimental and survey measures of time, risk, and social preferences: a review and some new results. *J. Dev. Econ.* 117:151–70

Chen RS, Boulding E, Schneider SH, eds. 1983. *Social Science Research and Climate Change*. Dordrecht, Neth.: D. Reidel

Corner A, Shaw C, Clarke J. 2018. *Principles for Effective Communication and Public Engagement on Climate Change: A Handbook for IPCC Authors*. Oxford, UK: Clim. Outreach

Curley SP. 2007. The application of Dempster-Shafer theory demonstrated with justification provided by legal evidence. *Judgm. Decis. Mak.* 2(5):257–76

Danileiko I, Lee MD. 2018. A model-based approach to the wisdom of the crowd in category learning. *Cogn. Sci.* 42(S3):861–83

Dasgupta I, Schulz E, Gershman SJ. 2017. Where do hypotheses come from? *Cogn. Psychol.* 96:1–25

Davis-Stober CP, Budescu DV, Dana J, Broomell SB. 2014. When is a crowd wise? *Decision* 1:79–101

Davis-Stober CP, Park S, Brown N, Regenwetter M. 2016. Reported violations of rationality may be aggregation artifacts. *PNAS* 113(33):E4761–63

Dawes RM, Faust D, Meehl P. 1989. Clinical versus actuarial judgment. *Science* 243:1668–74

Summary of research comparing predictive validity of expert judgments and statistical models.

Del Missier FD, Hansson P, Parker AM, Bruine de Bruin W, Nilsson L-G, Mantyla T. 2015. Unraveling the aging skein: disentangling sensory and cognitive predictors of age-related differences in decision making. *J. Behav. Decis. Mak.* 30(1):123–39

Dewitt B, Feeny D, Fischhoff B, Cella D, Hays RD, et al. 2018. Estimation of a preference-based summary score for the Patient-Reported Outcomes Measurement Information System: the PROMIS®-Preference (PROPr) scoring system. *Med. Decis. Mak.* 38(6):683–98

Dhami MK, Mandel DR, Mellers BA, Tetlock PE. 2015. Improving intelligence analysis with decision science. *Perspect. Psychol. Sci.* 10(6):753–57

Downs JS, Bruine de Bruin W, Fischhoff B. 2008. Patients' vaccination comprehension and decisions. *Vaccine* 26:1595–607

Downs JS, Murray PJ, Bruine de Bruin W, White JP, Palmgren C, Fischhoff B. 2004. Interactive video behavioral intervention to reduce adolescent females' STD risk: a randomized controlled trial. *Soc. Sci. Med.* 59(8):1561–72

Edwards W. 1954. A theory of decision making. *Psychol. Bull.* 51(4):380–417

Seminal article for the field, connecting analytical and behavioral approaches to decision making.

Edwards W. 1961. Behavioral decision theory. *Annu. Rev. Psychol.* 12:473–98

Edwards W, Lindman H, Savage LJ. 1963. Bayesian statistical inference for psychological research. *Psychol. Rev.* 70:193–242

Introduction to Bayesian inference, written for psychologists, providing normative standard for assessment of bias.

Elkind D. 1967. Egocentrism in adolescence. *Child Dev.* 38(4):1025–34

Engel L, Bansback N, Bryan S, Doyle-Waters MM, Whitehurst DGT. 2016. Exclusion criteria in national health state valuation studies: a systematic review. *Med. Decis. Mak.* 36(7):798–810

Ericsson KA, Simon HA. 1992. *Protocol Analysis*. Cambridge, MA: MIT Press

Ert A. 2018. Introduction to the concept of likelihood and its application. *Adv. Methods Pract. Psychol. Sci.* 1(1):60–69

Ert E, Erev I. 2013. On the descriptive value of loss aversion in decisions under risk: six clarifications. *Judgm. Decis. Mak.* 8(3):214–35

Farris C, Treat TA, Viken RJ. 2010. Alcohol alters men's perceptual and decisional processing of women's sexual interest. *J. Abnorm. Psychol.* 119(2):427–32

Farris C, Treat TA, Viken RJ, McFall RM. 2008. Sexual coercion and the misperception of sexual intent. *Clin. Psychol. Rev.* 28:48–66

FDA (Food Drug Adm.). 2018. *Benefit-Risk Assessment in Drug Regulatory Decision Making: Draft PDUFA VI Implementation Plan (FY2018–2022)*. Washington, DC: FDA

Fischhoff B. 1992. Giving advice: decision theory perspectives on sexual assault. *Am. Psychol.* 47:577–88

Fischhoff B. 1996. The real world: What good is it? *Organ. Behav. Hum. Decis. Process.* 65:232–48

Fischhoff B. 2008. Assessing adolescent decision-making competence. *Dev. Rev.* 28(1):12–28

Fischhoff B. 2015. The realities of risk-cost-benefit analysis. *Science* 350(6260):527

Introduction to behavioral aspects of formal analyses used in policy making.

Fischhoff B. 2017. Breaking ground for psychological science: the U.S. Food and Drug Administration. *Am. Psychol.* 72(2):118–25

Fischhoff B, Beyth-Marom R. 1983. Hypothesis evaluation from a Bayesian perspective. *Psychol. Rev.* 90:239–60

Fischhoff B, Bruine de Bruin W, Guvenc U, Caruso D, Brilliant L. 2006. Analyzing disaster risks and plans: an avian flu example. *J. Risk Uncertain.* 33:133–51

Fischhoff B, Kadvany J. 2011. *Risk: A Very Short Introduction*. Oxford, UK: Oxford Univ. Press

Fischhoff B, Lichtenstein S, Slovic P, Derby SL, Keeney RL. 1981. *Acceptable Risk*. New York: Cambridge Univ. Press

Gärdenfors P, Sahlin N-E, eds. 1988. *Decision, Probability, and Utility: Selected Readings*. Cambridge, UK: Cambridge Univ. Press

Girotto V, Gonzalez M. 2008. Children's understanding of posterior probability. *Cognition* 106:325–44

Glassman-Fox KT, Weber EU. 2016. What makes risks acceptable? Revisiting the 1978 psychological dimensions of risk. *J. Math. Psychol.* 75:157–69

Glöckner A, Hilbig BE, Henninger F, Fiedler S. 2016. The reversed description-experience gap: disentangling sources of presentation format effects in risky choice. *J. Exp. Psychol. Gen.* 145(5):486–508

Goldberg JH, Millstein S, Schwartz A, Halpern-Felsher B. 2009. Intertemporal tradeoffs: perceiving the risk in the benefits of marijuana in a prospective study of adolescents and young adults. *Med. Dec. Mak.* 29:182–92
Gonzalez C, Mehlhorn K. 2016. Framing from experience: cognitive processes and predictions of risky choice. *Cogn. Sci.* 40:1163–91
Grossmann I. 2017. Wisdom in context. *Perspect. Psychol. Sci.* 12(2):233–57
Hadar L, Fox CR. 2009. Information asymmetries in decisions from description versus decisions from experience. *Judgm. Decis. Mak.* 4:317–25
Hall CC, Zhao J, Shafir E. 2014. Self-affirmation among the poor: cognitive and behavioral implications. *Psychol. Sci.* 25(2):619–25
Hertwig R, Barron G, Weber ER, Erev I. 2004. Decisions from experience and the effect of rare events in risky choice. *Psychol. Sci.* 15(8):534–39
Hess TM, Strough J, Lockenhoff CE, eds. 2015. *Aging and Decision Making*. New York: Academic
Heuer RJ. 1999. *Psychology of Intelligence Analysis*. Washington, DC: Cent. Study Intell., Cent. Intell. Agency
Ho EH, Budescu DV, Dhami MK, Mandel DR. 2015. Improving the communication of uncertainty in climate science and intelligence analysis. *Behav. Sci. Policy* 1(2):53–66
Huber GP. 1983. Cognitive style as a basis for MIS and DSS designs: much ADO about nothing? *Manag. Sci.* 29:567–79
Huber J, Payne JW, Puto CP. 2014. Let's be honest about the attraction effect. *J. Mark. Res.* 51(8):520–25
Jaccard J. 2012. The reasoned action model: directions for future research. *Ann. Am. Acad. Political Soc. Sci.* 640:58–80
Jackson SA, Kleitman S, Stankov L, Howie P. 2017. Individual differences in decision making depend on cognitive abilities, monitoring, and control. *J. Behav. Decis. Mak.* 30(2):209–23
Jacobs JE, Klaczynski P, eds. 2005. *The Development of Judgment and Decision Making in Children and Adults*. Mahwah, NJ: Lawrence Erlbaum
Kahneman C. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux
Kahneman D, Klein G. 2009. Conditions for intuitive expertise. *Am. Psychol.* 64:515–26
Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47(2):263–92
Kaplan EH. 2011. Operations research and intelligence analysis. In *Intelligence Analysis: Behavioral and Social Science Foundations*, ed. Natl. Res. Counc, pp. 31–56. Washington, DC: Natl. Acad. Press
Karelaia N, Hogarth R. 2008. Determinants of linear judgment: a meta-analysis of lens model studies. *Psychol. Bull.* 134(3):404–26
Kellen D, Pachur T, Hertwig R. 2016. How (in)variant are subjective representations of described and experienced risk and rewards? *Cognition* 157:126–38
Kent S. 1964. Words of estimative probability. *Stud. Intell.* 8(4):49–65
Kyburg HE, Smokler HE, eds. 1964. *Studies in Subjective Probability*. New York: Wiley
Lanir Z, Kahneman D. 2006. An experiment in decision analysis in Israel in 1975. *Stud. Intell.* 50(4):11–19
Lerner JS. 2019. Decision science meets national security: a personal perspective. *Perspect. Psychol. Sci.* 14(1):96–100
Lerner JS, Li Y, Valdesolo P, Kassam KS. 2015. Emotion and decision making. *Annu. Rev. Psychol.* 66:799–823
Levy BR, Slade MD, Pietrzak RH, Ferucci L. 2018. Positive age beliefs protect against dementia even among elders with high-risk gene. *PLOS ONE* 13(2):e0191004
Lichtenstein S, Fischhoff B. 1980. Training for calibration. *Org. Behav. Hum. Perform.* 26:149–71
Lichtenstein S, Slovic P, eds. 2006. *Construction of Preferences*. London: Earthscan
Lieder F, Griffiths TL, Hsu M. 2018. Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychol. Rev.* 125(1):1–32
Lynn SK, Barrett LF. 2014. "Utilizing" signal detection theory. *Psychol. Sci.* 25(9):1663–73
Mannes AE, Soll JB, Larrick RP. 2014. The wisdom of select crowds. *J. Pers. Soc. Psychol.* 107(2):276–99
May R. 1991. Overconfidence in overconfidence. In *Progress in Decision Utility and Risk Theory*, ed. A Chikán, pp. 67–76. Dordrecht, Neth.: Springer
Strong statement regarding the need for, and benefits of, diversity in all aspects of behavioral research.
Medin D, Ojalehto B, Marin A, Berg M. 2017. Systems of (non-)diversity. *Nat. Hum. Behav.* 1:0088

Merkle EC, Steyvers M. 2013. Choosing a strictly proper scoring rule. *Decis. Anal.* 10(4):292–304
Meyerhoff J, Liebe U. 2010. Determinants of protest responses in environmental valuation: a meta-study. *Ecol. Econ.* 70:366–74
Milkman KL, Chugh D, Bazerman MH. 2009. How can decision making be improved? *Perspect. Psychol. Sci.* 4:379–83
Mohan D, Farris C, Fischhoff B, Rosengart MR, Angus D, et al. 2017. Testing the efficacy of a video game versus a traditional education program at improving physician decision making in trauma triage: a randomized controlled trial. *BMJ* 359:j5416
Mohan D, Rosengart MR, Farris C, Fischhoff B, Angus DC, Barnato AE. 2012. Sources of non-compliance with clinical practice guidelines in trauma triage: a decision science study. *Implement. Sci.* 7:103
Moore DA, Swift SA, Minster A, Mellers BA, Ungar LH, et al. 2017. Confidence calibration in a multiyear geopolitical forecasting competition. *Manag. Sci.* 63(11):3552–65
Morgan MG. 2017. *Theory and Practice of Policy Analysis*. New York: Cambridge Univ. Press
Mullainathan S, Shafir E. 2013. *Scarcity*. New York: Picador
Murphy AH, Winkler RE. 1974. Probability forecasts. *Bull. Am. Meteorol. Soc.* 55:1449–53
Natl. Res. Counc. 2011. *Intelligence Analysis for Tomorrow*. Washington, DC: Natl. Acad. Press
Natl. Sci. Board. 2014. *Science and Engineering Indicators 2014 (NSB 14-01)*. Arlington, VA: Natl. Sci. Found. **https://www.nsf.gov/statistics/seind14/content/etc/nsb1401.pdf**
O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, et al. 2006. *Uncertain Judgements: Eliciting Expert Probabilities*. Chichester, UK: Wiley
Oppenheimer DM, Kelso E. 2015. Information processing as a paradigm for decision making. *Annu. Rev. Psychol.* 66:277–94
Ott. Hosp. 2019. Patient decision aids. *The Ottawa Hospital Research Institute*. **https://decisionaid.ohri.ca/index.html**
Parker AM, Bruine de Bruin W, Fischhoff B, Weller J. 2018. Robustness of decision-making competence: evidence from two measures and an 11-year longitudinal study. *J. Behav. Decis. Mak.* 31:380–91
Parker AM, Fischhoff B. 2005. Decision-making competence: external validity through an individual-differences approach. *J. Behav. Decis. Mak.* 18:1–27
Pashler H, McDaniels M, Rohrer D, Bjork R. 2009. Learning styles: concepts and evidence. *Psychol. Sci. Public Interest* 9(3):105–19
Payne JW, Bettman JR, Johnson EJ. 1993. *The Adaptive Decision Maker*. New York: Cambridge Univ. Press
Pedroni A, Frey R, Bruhin A, Dutilh G, Hertwig R. 2018. The risk elicitation puzzle. *Nat. Hum. Behav.* 1:803–9
Podsakoff PM, MacKenzie SB, Podsakoff NP. 2012. Sources of method bias in social science research and recommendations on how to control it. *Annu. Rev. Psychol.* 63(1):539–69
Raiffa H. 1968. *Decision Analysis*. Reading, MA: Addison-Wesley
Rapoport A, Wallsten TS. 1972. Individual decision behavior. *Annu. Rev. Psychol.* 23:131–76
Regenwetter M, Robinson MM. 2017. The construct-behavior gap in behavioral decision research: a challenge beyond replicability. *Psychol. Rev.* 124(5):533–50
Reyna VF. 2012. A new intuitionism: meaning, memory, and development in fuzzy-trace theory. *Judgm. Decis. Mak.* 7(3):332–59
Reyna VF, Chapman SB, Dougherty MR, Confrey J, eds. 2012. *The Adolescent Brain*. Washington, DC: Am. Psychol. Assoc.
Reyna VF, Mills BA. 2014. Theoretically motivated interventions for reducing sexual risk taking in adolescence: a randomized controlled experiment applying fuzzy-trace theory. *J. Exp. Psychol. Gen.* 143(4):1627–48
Roper v. Simmons, 543 U.S. 03-633 (2005)
Ross M, Grossmann I, Schryer E. 2014. Contrary to psychological and popular opinion, there is no compelling evidence that older adults are disproportionately victimized by consumer fraud. *Perspect. Psychol. Sci.* 9:427–42
Salekin RT. 2015. *Forensic Evaluation and Treatment of Juveniles*. Washington, DC: Am. Psychol. Assoc.
Scherer LD, Burke JF, Zikmund-Fisher BJ, Caverly T, Kullgren J, et al. 2016. Development of the Medical Maximizer-Minimizer Scale. *Health Psychol.* 35:1276–87

Accessible introduction to policy analysis, highlighting the centrality of behavioral decision research to the field.

Schulte-Mecklenbeck M, Johnson JG, Bockenholt U, Goldstein DG, Russo JE, et al. 2017. Process-tracing methods in decision making. *Curr. Dir. Psychol. Sci.* 26(5):442–50

Schum DA. 1994. *The Evidential Foundations of Probabilistic Reasoning*. New York: Wiley

Schwartz A, Bergus G. 2008. *Medical Decision Making*. New York: Cambridge Univ. Press

Schwartz LM, Woloshin S. 2013. The Drug Facts Box: improving the communication of prescription drug information. *PNAS* 110(Suppl. 3):14069–74

Shafer G, Tversky A. 1986. Languages and designs for probability judgment. *Cogn. Sci.* 9:309–39

Slovic P. 1964. Assessment of risk taking. *Psychol. Bull.* 61(3):220–23

Slovic P. 2001. *Feeling of Risk*. London: Earthscan

Slovic P, Fischhoff B, Lichtenstein S. 1977. Behavioral decision theory. *Annu. Rev. Psychol.* 28:1–39

Slovic P, Lichtenstein S. 1971. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Org. Behav. Human Perform.* 6:649–744

Comprehensive review of early behavioral decision research, presaging future directions.

Slovic S, Slovic P, eds. 2015. *Numbers and Nerves: Information, Emotion, and Meaning in a World of Data*. Corvallis: Or. State Univ. Press

Sniezek JA, Buckley T. 1991. Confidence depends on level of aggregation. *J. Behav. Decis. Mak.* 4:263–72

Steinberg L, Cauffman E, Woolard J, Graham S, Banich M. 2009. Are adolescents less mature than adults? Minors' access to abortion, the juvenile death penalty, and the alleged APA "flip-flop." *Am. Psychol.* 64(7):583–94

Stewart N, Chater N, Brown GD. 2006. Decision by sampling. *Cogn. Psychol.* 53(1):1–26

Stewart N, Reimers S, Harris AJL. 2015. On the origin of utility, weighting, and discounting functions: how they get their shapes and how to change their shapes. *Manag. Sci.* 61(3):687–705

Svenson O. 1979. Process descriptions of decision making. *Org. Behav. Hum. Perform.* 23:86–112

Swets JA, Dawes RM, Monahan J. 2000. Psychological science can improve diagnostic decisions. *Psychol. Sci. Public Interest* 1(1):1–26

Introduction to diagnostic decisions, showing diverse applications of behaviorally informed signal detection theory research.

Szaszi B, Palinkas A, Palfi B, Szollosi A, Aczel B. 2017. A systematic scoping review of the choice architecture movement: toward understanding when and why nudges work. *J. Behav. Decis. Mak.* 31(7):355–66

Tarter R, Vanyukov M. 2001. Introduction: theoretical and operational framework for research into the etiology of substance use disorder. *J. Child Adolesc. Subst. Abuse* 10(1):1–12

Thaler R, Sunstein C. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale Univ. Press

Tomassini LA, Solomon I, Romney MB, Krogstad JL. 1982. Calibration of auditors' probabilistic judgments: some empirical evidence. *Org. Behav. Hum. Perform.* 30:391–406

Tong J, Feiler D. 2017. A behavioral model of forecasting: naïve statistics on mental models. *Manag. Sci.* 63(11):3609–27

Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185(4157):1124–31

Seminal article for heuristics and biases approaches to judgment and decision making.

Tversky A, Kahneman D. 1983. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90:293–315

Tversky A, Kahneman D. 1992. Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5(4):297–323

Üklümen G, Fox CR, Malle BF. 2016. Two dimensions of subjective uncertainty: clues from natural language. *J. Exp. Psychol. Gen.* 145(10):1280–97

van der Linden S, Leiserowitz A, Maibach E. 2018. Scientific agreement can neutralize the politicization of facts. *Nat. Human Behav.* 2:2–3

von Neumann J, Morgenstern O. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press

von Winterfeldt D. 2013. Bridging the gap between science and decision making. *PNAS* 110:14055–61

von Winterfeldt D, Edwards W. 1986. *Decision Analysis and Behavioral Research*. New York: Cambridge Univ. Press

Walasek L, Mullett TL, Stewart N. 2018. *A meta-analysis of loss aversion in risky contexts*. Work. Pap., Univ. Warwick, Coventry, UK

Walters DJ, Ferbach PM, Fox CR, Sloman SA. 2017. Known unknowns: a critical determinant of confidence and calibration. *Manag. Sci.* 63(12):4298–307

Weber EU, Blais A-R, Betz N. 2002. A domain-specific risk-attitude scale. *J. Behav. Decis. Mak.* 15:263–90

Weber EU, Johnson EJ. 2009. Mindful judgment and decision making. *Annu. Rev. Psychol.* 60:53–86

Wickelgren W. 1977. Speed-accuracy trade-off and information processing dynamics. *Acta Psychol*. 41:67–85

Wong-Parodi G, Krishnamurti T, Davis AL, Schwartz D, Fischhoff B. 2016. Integrating social science in climate and energy solutions: a decision science approach. *Nat. Clim. Change* 6:563–69

Applications of decision science to diverse climate- and energy-related decisions.

Zikmund-Fisher B. 2019. Helping people know whether measurements have good or bad implications: increasing the evaluability of health and science data communications. *Policy Insights Behav. Brain Sci.* 6(1):29–37