Advanced Review

Judgment and decision making

Baruch Fischhoff*

The study of judgment and decision making entails three interrelated forms of research: (1) *normative analysis*, identifying the best courses of action, given decision makers' values; (2) *descriptive studies*, examining actual behavior in terms comparable to the normative analyses; and (3) *prescriptive interventions*, helping individuals to make better choices, bridging the gap between the normative ideal and the descriptive reality. The research is grounded in analytical foundations shared by economics, psychology, philosophy, and management science. Those foundations provide a framework for accommodating affective and social factors that shape and complement the cognitive processes of decision making. The decision sciences have grown through applications requiring collaboration with subject matter experts, familiar with the substance of the choices and the opportunities for interventions. Over the past half century, the field has shifted its emphasis from predicting choices, which can be successful without theoretical insight, to understanding the processes shaping them. Those processes are often revealed through biases that suggest non-normative processes. The practical importance of these biases depends on the sensitivity of specific decisions and the support that individuals have in making them. As a result, the field offers no simple summary of individuals' competence as decision makers, but a suite of theories and methods suited to capturing these sensitivities. *WIREs Cogn Sci* 2010 1 000–0000

Decisions are easy when decision makers know what they want and what they will get, making choices from a set of well-defined options. Such decisions could be equally easy, but reach different conclusions, for people who see the facts similarly, but have different goals, or for people who have the same values but see the facts differently, or for people who disagree about both facts and values.

Decision making can become more difficult when there is uncertainty about either what will happen or what one wants to happen. Some decisions are so sensitive to estimates of fact or value that it pays to invest in learning, before acting. Other decisions will work out just as well, for any plausible estimates.

Thus, any account of decision-making processes must consider both the decisions and the individuals making them. The field of *behavioral decision research* provides such accounts. It entails three forms of research: (1) *normative*, identifying the best possible choice, given the state of the world and decision makers' values; (2) *descriptive*, characterizing how individuals make decisions, in terms comparable to the normative standard; and (3) *prescriptive*, attempting to close the gap between the normative ideal and the descriptive reality.

Although they can be described as an orderly progression, these three forms of research are deeply interrelated. Descriptive research is needed to reveal the facts and values that normative analysis must consider. Prescriptive interventions are needed to assess whether descriptive accounts provide the insight needed to improve decision making. Normative analyses are needed to understand the facts that decision makers must grasp and the practical implications of holding different values. Thus, understanding choices requires an iterative process, cycling through the three stages. This chapter follows the evolution of theory and method for seeking that understanding.

BEHAVIORAL DECISION RESEARCH

Behavioral decision research emerged from normative models of decision making developed by philosophers,

*Correspondence to: baruch@cmu.edu

Department of Social and Decision Sciences,• and Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, 15213-3890, USA

DOI: 10.1002/wcs.65

mathematicians, and economists.[1,2] These models describe how to determine the best possible course of action, given what individuals believe about the world and what they want from it. Individuals who follow these rules are said to be *rational*. Their choices are *optimal*, if they are well informed about the world and about their own values. Although normative models take strong positions on *how* decisions should be made, they are mute regarding *what* options, facts, and values should be considered. As a result, they require the empirical content provided by descriptive and prescriptive research to be anything but formalisms.

Comparability with normative analysis imposes important constraints on descriptive and prescriptive research. They cannot begin without first examining the world from decision makers' perspective. They cannot criticize choices without asking whether they might be rational, given what people want and believe. They cannot assess the importance of imperfections in decision-making processes, without embedding them in normative analyses, showing their practical implications. Imperfections can be theoretically informative without mattering much. Indeed, nonrational processes may survive because they have too little practical significance to provide the sharp negative feedback sometimes needed to change behavior.

Psychology progresses, in part, by applying what Berkeley and Humphreys[3] call the 'bias heuristic', identifying departures from normative standards.[4] However, unless those standards are well defined, vaguely similar biases may proliferate. Different biases might share a common name (e.g., confirmation bias); the same bias might have different names (e.g., saliency, availability), impeding scientific progress.[5] Indeed, as discussed next, a major advance in early behavioral decision research was discovering that seemingly different theories were often indistinguishable.

CLINICAL JUDGMENT

World War II was a turning point for psychology, which showed its ability to assess efficiently the skills and problems of masses of individuals. After the war, attention turned to the effectiveness of those efficient assessments. These studies of *clinical judgment* quickly spread to topics as diverse as how psychologists decide whether clients are 'neurotic' or 'psychotic', radiologists sort ulcer X-rays into 'benign' or 'malignant', bank officers classify loans as 'nonperforming', and brokers weigh stocks' prospects.[6–8]

Conducting studies of clinical judgment is straightforward: Collect many judgments of cases described on a common set of possibly relevant *cues*. Use statistical methods (e.g., multiple regression) to predict those judgments from the cues. For example, Dawes[9] studied University of Oregon Psychology Department graduate admission committee evaluations of 384 applicants. Although applicants' files had many cues (e.g., letters of recommendation, full transcripts), the committee's ratings could be predicted well from just three: Graduate Record Examination (GRE) score, undergraduate grade point average (GPA), and quality of undergraduate institution (QI):

$$0.0032\ \mathrm{GRE} + 1.02\ \mathrm{GPA} + 0.0791\ \mathrm{QI} \qquad (1)$$

This study illustrates four frequently replicated patterns:[9,10] (1) A simple model predicts a seemingly complex process. (2) Judges describe using very different strategies than that 'captured' in the model. For example, committee members claimed that they considered more cues and used these three in more nuanced ways than just weighting and adding. (3) Even simpler models, replacing regression weights with unit weights on normalized variables, predict equally well. (4) Simple models predict the actual criterion (graduate school success) well.

There are at least three reasons why simple models predict surprisingly well. One is that people have difficulty introspecting into their own decision making.[11,12] A second is that people have difficulty executing complex strategies reliably, so that only simple patterns appear consistently. The third is that simple linear models can predict well without capturing the underlying processes,[9,13] as long as they use reliably measured correlates of the variables that actually affect decision making.

This good news for predictive research is bad news for explanatory research. Models using different variables, implying different processes, often predict equally well. As a result, regression weights need not capture how decisions are made. In many applications, good prediction suffices. For example, the health belief model[14,15] provides a structured way to identify variables correlated with health-related choices. Its application would, however, be misguided, if the weights on those variables were taken as reflecting how individuals think.[16]

For analogous reasons, behavioral decision researchers typically avoid the *revealed preference* analyses that are a staple of economics research.[17,18] For goods traded in efficient markets, prices show rational decision makers' values. If goods are characterized on common attributes, regression weights show those attributes' usefulness as predictors. For example, house prices might be predicted from their size, condition, age, school district, commuting distance, construction, and so on. Unfortunately, when predictors are correlated, regression weights can be unstable, complicating their interpretation as measures of importance.

One strategy for undoing these confounds is generating stimuli with uncorrelated cues. For example, one might create hypothetical graduate school candidates, using all possible combinations of GRE, GPA, and QI. A drawback to this *ANOVA design* is violating behavioral decision research's commitment to *probabilistic functionalism*,[4,19] the view that behavior is shaped by naturally occurring correlations among uncertain cues. Stimuli that violate these relationships lack ecological validity and require unnatural behavior, such as evaluations of implausible cue combinations (e.g., low GRE, high QI). An ANOVA design also gives equal weight to all cue combinations, however, common or possible. Moreover, as with any design that presents many stimuli with a transparent cue structure, respondents may either lose focus (producing unreliable judgments) or improvise a mechanical response strategy (producing reliable, but unnatural judgments).

How people respond to novel tasks (e.g., grad candidates with low GRE and high QI) can be revealing. However, because *importance* is inherently context dependent, artificial contexts produce artificial importance weights. For example, although money is generally relevant to consumer decision making, other factors may dominate choices among similarly priced options. One possible reason why Eq. (1) did not include the variable 'strength of letters of recommendation' is that candidates had similarly strong letters, written by faculty advisors who sell their students similarly. (QI should capture the reputations of those letter writers.)

As it discovered these limits to the explanatory value of predictive models, behavioral decision research shifted its focus from *what* choices people make to *how* they make them. As a result, studies describe decision-making processes that *can* come into play, as revealed by tasks with which it is relatively clear how a process would express itself. Applied researchers must then determine which of the possible processes are evoked by a specific decision.

SUBJECTIVE EXPECTED UTILITY

The normative analysis underlying behavioral decision research is founded on *expected utility theory*, classically codified by von Neumann and Morgenstern.[20] Its basic logic is straightforward: List the possible action options. For each option, enumerate its possible outcomes. For each such outcome, assess the value, or *utility*, of it happening. Assess the probability of its occurrence should each option be selected. Compute the *expected utility* of each option by multiplying the utility and probability of each outcome, should it be undertaken, then summing across outcomes. Choose the action with the greatest expected utility. When the probabilities reflect decision makers' beliefs, rather than scientific knowledge, the calculation produces *subjective expected utility*.[21] (As discussed below, some scholars view all probabilities as subjective.)

Descriptive research can look at how people undertake each element of this process: assessing the probabilities of possible outcomes, evaluating their utility (should they occur), and combining probabilities and utilities to identify the best option. The decisions can range from completely described and static to incompletely described and dynamic. Normative analyses exist for many kinds of decision.[22]

Individuals' performance on these tasks can be evaluated by *correspondence* or *coherence* tests. Correspondence tests ask how accurate their answers are. For example, how well can they predict whether they will graduate college or enjoy their major? Coherence tests ask how consistent responses are. For example, are probability judgments for event at least as large as those for subset ($p[A] \geq p[A \cap B]$)? Are outcomes equally valued, when described in formally equivalent ways (e.g., succeeding vs. not failing).

PREDICTING OUTCOMES

Studies of how well people predict uncertain events have produced seemingly contradictory results. Sometimes, people do quite well; sometimes, quite poorly. To a first approximation, the difference depends on whether the task requires counting or inferences. With counting studies, the evidence is all of one type; with inference studies, the evidence is of different types. A counting study might display stimuli drawn randomly from a hidden population, then elicit estimates of a summary statistic (e.g., mean, range).[23] An inference study might require integrating *base-rate* evidence about what usually happens, with *individual information* about a specific case.

Counting tasks take advantage of individuals' ability to estimate the relative frequency of events that they observe—even without preparing to do so. For example, after producing rhymes for a set of words, people can estimate the number beginning with different letters.[24] Indeed, encoding frequencies has been called an automatic cognitive function, with research

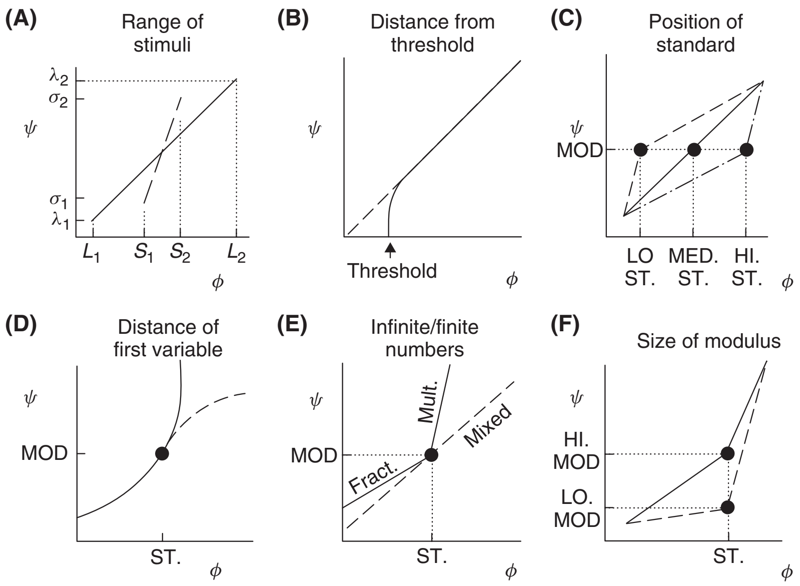

FIGURE 1 | Six 'laws of the new psychophysics', depicting the influence of experimental design on the numerical response used to describe the psychological state (ψ) equivalent to a physical stimulus (ϕ). (A) A narrower stimulus range ($S1, S2$) will use a proportionately larger portion of the response range than would the same stimuli, when embedded in a larger response range ($L1, L2$). (B) The effects of assumptions regarding the treatment of stimuli below the threshold of perception or evaluation. (C) The effects of where a standard stimulus falls in the response range, after it has been assigned a numerical valuation (or modulus). (D) The effects of where the first judged stimulus is relative to the standard. (E) The effects of using fractional or integer response values, for stimuli smaller than the standard. (F) The reverse effects where a modulus value, for a given standard stimulus, falls within the response range. Source: Fischhoff (2005).[17]

focusing on whether it relies on *tokens*, records of individual observations, or on *types*, category representatives reinforced with each observation.[25]

Assuming that individuals trust their frequency-encoding ability, Tversky and Kahneman[26] proposed the *availability heuristic*, whereby individuals estimate an event's probability by their ability to retrieve instances (tokens) or imagine them (types). Reliance on availability produces biased judgments when the observed events are an unrepresentative sample—and individuals cannot correct for the sampling bias. Researchers have identified many other possible biases, arising from reliance on judgmental heuristics. The strength of any claim of bias depends on the strength of the normative analysis.[27,28] The usefulness of any heuristic depends on how well its application can be predicted (e.g., how memory is searched, for examples).

Inference studies tap individuals' lack of intuition and training for combining different kinds of evidence. Here, the normative standard has been the Bayesian approach to hypothesis evaluation.[5,29] Bayes theorem is an uncontroversial part of probability theory. Bayesian inference is more controversial, because it treats probabilities as subjective, thereby allowing inferences that combine diverse kinds of evidence.[30] *Frequentistic* probabilities require evidence of a single kind (e.g., coin flips, weather records). Subjective judgments are only probabilities if they pass coherence tests. Thus, probabilities are not just any assertion of belief.

A widely studied inferential bias is the 'base-rate fallacy'. Attributed to reliance on the representativeness heuristic,[26] it involves allowing even weak information about specific cases to outweigh knowledge of what generally happens (the base rate). Inadequately regressing judgments is the same bias with continuous variables. Absent strong information about specific cases, one should predict the mean of a distribution.

To avoid artifactual sources of bias,[31] behavioral decision research draws on the century-plus of psychophysics research into factors affecting quantitative judgments.[32–34] For example, because people avoid decimals, they are more likely to overestimate small risks in studies eliciting percentages (e.g., 0.1%) than in studies eliciting odds (e.g., 1 in 1000).[33] Knowing that, researchers can choose the method best suited to their question and avoid unsupported claims.

Figure 1 depicts six such design features, critical to eliciting numbers. For example, Figure 1A shows that stimuli [S_1, S_2] elicit less of the response range when embedded in a larger range [L_1, L_2]. Figure 1C shows how values assigned to larger stimuli are cramped if the initial (standard) stimulus is large, relative to others in the set. Such effects occur because respondents must translate their perceptions into the investigators' terms. Where those terms are unnatural, respondents rely on *response preferences*.[35] For example, they try to use the entire response scale; they look for patterns in randomly ordered stimuli; they deduce an expected level of precision, such as what trade-off to make between speed and accuracy.

Ignoring response preferences leads to misinterpreting judgments. For example, subjects produced much higher estimates of annual US death toll, from 41 causes, when they received a high *anchor* (50,000 motor vehicles deaths), rather than a low anchor (1000 accidental electrocutions). Low frequencies were greatly overestimated with the high anchor, but not with the low one. Estimates of relative frequency were similar; however, the question was asked, suggesting robust risk perceptions, whose translation into numerical judgments was method dependent. The estimates were also biased in ways consistent with relying on the availability heuristic (e.g., homicides were overestimated relative to the, less reported, suicides).[36]

ELICITING VALUES

There are two streams of research into how people form preferences.[17] One follows psychophysics, treating the intensity of preferences like the intensity of physical experiences. The second follows the precepts of *decision analysis*, a consulting process designed to help individuals follow decision theory's normative model.[21,22]

Research in the psychophysical stream has individuals report their feelings directly, perhaps with a rating scale ora judgment willingness-to-pay for a good. Attitude research is the archetype of this paradigm.

The correspondence test for psychophysical research asks how well elicited values predict behavior. Some attitude researchers hold that a fair test must elicit attitudes that are directly comparable to the target behavior.[37] For example, many behaviors could follow endorsement of ‘my faith is very important to me’. Stronger predictions follow from ‘daily prayer with like-minded worshipers is very important to me’. Even stronger predictions follow from specifying the form of worship. At the extreme, these judgments become statements of intention, rather than attitudes, representing general values. As such, their validity depends on how well people can predict their own experiences.[38]

The coherence standard for psychophysical judgments is *construct validity*. Expressed values should be sensitive to relevant changes in questions and insensitive to irrelevant ones. Applying this standard requires independently assessing relevance. For example, assuming that more is better, *scope tests* ask whether people put higher values on larger quantities of a good. Scope insensitive judgments represent incoherent preferences—except for individuals who feel that there can be too much of a good thing (e.g., rich food, conspicuous consumption). An ‘inside view’ on individuals' basic values is needed to evaluate the coherence of their preferences.

Research in the decision analysis stream assumes that people cannot know what they want, in all possible situations. Rather, they must *construct* specific preferences from more basic values. In making these inferences, people may seek cues in a world that might be helpful, indifferent, or manipulative. The better people understand the factors shaping their inferences, the better chance they have of figuring out what they want.[39,40] Decision analysis structures that process. Its measurement is *reactive*, in the sense of changing people in the process of trying to help them discover their preferences. If successful, it deepens individuals' understanding of themselves.

Correspondence tests for constructed preferences compare elicited values with those that emerge from similar real-world processes. Thus, an intensive electoral campaign might be the standard for a study eliciting candidate preferences.[41,42] Intensive consultation medical experience might be the standard for preferences elicited with a medical decision aid.[43,44] Coherence tests for constructed preferences ask whether the elicitation session has included all perspectives that individuals might want to consider, while avoiding ones that would apply irrelevant influences.

Identifying the factors influencing behavior is, of course, psychology's central challenge. To study theoretically relevant factors, researchers must control irrelevant ones. Understanding these processes is an ongoing enterprise, which McGuire [45] depicted as turning ‘artifacts into main effects’, worthy of independent investigation. Table 1 assembles parts of this history, in terms of the four essential elements of any behavior: the organism, the stimulus being evaluated, the response mode for expressing preferences, and potentially distracting contexts.[46] In terms of correspondence tests, these are all factors that could undermine the match between the conditions in which values are measured by researchers and expressed in life. In terms of coherence tests, these are all factors whose effects on expressed values could be compared with independent assessments of their relevance. That is, do changes in these factors affect valuations when, and only when, they should make a difference? Given the sheer number of potentially relevant factors, value elicitation requires broad understanding of behavioral science.

MAKING DECISIONS

Non-normative Theories

Knowing the limits to the theoretical insights possible with predictive models, applied in complex settings, behavioral decision researchers have focused on processes observed most clearly under experimental conditions. The robustness of observations in the lab is tested by varying those conditions (e.g., increasing economic incentives for good performance, changing

TABLE 1 | From Artifact to Main Effect

Liability in judgment due to	Led to
Organism	
Inattention, laziness, fatigue, habituation, learning, maturation, physiological limitations, natural rhythms, experience with related tasks	Repeated measures
	Professional subjects
	Stochastic response models
	Psychophysiology
	Proactive and retroactive inhibition research
Stimulus presentation	
Homogeneity of alternatives, similarity of successive alternatives (especially first and second), speed of presentation, amount of information, range of alternatives, place in range of first alternative, distance from threshold, order of presentation, areal extent, ascending or descending series	Classic psychophysical methods
	The new psychophysics
	Attention research
	Range-frequency theory
	Order-effects research
	Regression effects
	Anticipation
Response mode	
Stimulus-response compatibility, naturalness of response, set, number of categories, halo effects, anchoring, very small numbers, response category labeling, use of end points	Ergonomics research
	Set research
	Attitude measurement
	Assessment techniques
	Contrasts of between- and within-subject design
	Response-bias research
	Use of blank trials
'Irrelevant' context effects	
Perceptual defenses, experimenter cues, social pressures, presuppositions, implicit payoffs, social desirability, confusing instructions response norms, response priming, stereotypic responses, second-guessing	New look in perception
	Verbal conditioning
	Experimenter demand
	Signal-detection theory
	Social pressure, comparison, and facilitation research

Source: Fischhoff et al. (1980).[46]

information displays) and by identifying real-world analogs, in which a theoretically interesting process might play a practical role.

Foremost among these models is Kahneman and Tversky's[47] *prospect theory*. Its initial formulation identified several utility theory assumptions that were implausible psychologically. One is that people evaluate expected outcomes in terms of changes in their *net asset position*, namely, everything they have in the world. However, people are actually highly sensitive to changes and tend to forget the big picture—as witnessed in reminders to 'count your blessings'.[48] A second psychologically implausible assumption is that numerically equivalent changes in probabilities are equally important. However, the psychophysics of probability weighting places a premium on changes that lead to certain outcomes (e.g., from 90% to 100%) compared to mid-range changes (e.g., from 30% to 40%). A third such assumption is that people get increasing averse, as losses mount up, whereas psychology finds them increasingly apathetic.

One widely studied corollary of these principles is the *status quo bias*. It reflects how easily *reference points* can be shifted, varying how changes are viewed. For example, organ donation rates are much higher when drivers must opt out, when getting their drivers licenses, compared to when they must opt in.[49,50] Opting in makes surrendering organs seem like a loss, hence aversive. That formulation also suggests a social norm of organ donation and perhaps even a weaker right to refusal.

Any behaviorally realistic approach to decision making must accommodate the limits to cognitive computational capacity. Prospect theory accepts utility theory's cognitively implausible calculation of expected values. However, it uses more intuitively plausible elements and, as a linear model, is relatively robust to misestimating its parameters. Applying the theory requires identifying its elements with real-world equivalents, such as the reference points that decision makers use when assessing changes.[51] Fuzzy-trace theory[52] studies the processes by which individuals master the gist of recurrent decisions. Approaches building on the classic work of Herbert Simon[53] have examined individuals' ability to match simple decision-making heuristics to choices that would, otherwise, be unduly complex.[54] Query theory,[55] support theory,[56] and others [57] formalize the notion of weighting retrieved beliefs, embodied in the availability heuristic.

Emotions

Normative analyses can accommodate emotions as valued outcomes, such as the utility of being happy or the disutility of being fearful. For example, there are formal methods for incorporating such 'psychological' outcomes, in analyses of risk decisions.[17,31] Descriptive research can accommodate emotions in terms of their effects on each element of decision making (defining options, predicting events, assessing personal values, integrating beliefs and values). For example, *cognitive appraisal theory*[58] predicts that anger increases the perceived probability of overcoming problems. In a field test with a nationally representative US sample, Lerner et al.[59] found that respondents were about 6% more optimistic, regarding their vulnerability to terror-related events, after an anger induction than after a fear induction. Prescriptive research can accommodate emotions by helping people to getting the right mix for particular choices.[60,61] For example, formal analyses might be used cautiously when they 'anaesthetize' moral feeling;[62] decision aids for adolescents have focused on controlling emotions.[63]

The importance of emotion effects depends on their size. A 6% shift might tip a close decision, but not a clear-cut one. von Winterfeldt and Edwards[21] showed, mathematically, that decisions with continuous options (e.g., invest $X) are often insensitive to changes in input variables (i.e., probabilities, values). Thorngate [64] used simulations to examine the sensitivity of stylized decisions to errors due to imperfect heuristics, an approach that others have pursued extensively.[65,66]

Decision-making Competence (DMC)

The fundamental premise of experimental decision research is that people who master the skills that it studies make better real-world decisions.[67] Table 2 presents results from a study evaluating the external validity of seven experimental tasks, chosen to span the space of cognitive decision-making competencies.[68] Respondents were 110 18- to 19-year-old males in a longitudinal study involving extensive assessments beginning at age 10. DMC scores, extracted from a factor analysis of performance on the seven tasks, showed good test–retest reliability, as did scores on an adult version.[69]

The first section shows positive correlations between DMC and standard measures of verbal and fluid intelligence (Vocabulary and ECF, respectively). The second section shows positive correlations between DMC and four 'constructive' cognitive styles. The third section shows negative correlations between DMC and several important risk behaviors. The fourth section shows that DMC is higher for teens coming from low-risk (LAR) families, higher socioeconomic status (SES) families, and more positive peer environments. (The negative correlation with social support may reflect low DMC teens' greater gang membership.) Most correlations remained statistically significant after partialing out the two intelligence measures.

These results support the construct validity of DMC as a measure of decision-making skills that both cause and reflect important aspects of teens' lives. For example, teens with higher DMC come from families that might both model and reward good decision making. Bruine de Bruin et al.[69] found similar correlations between adult DMC and scores on a psychometrically validated Decision Outcome Inventory, eliciting self-reports of outcomes suggesting poor decisions, varying in severity (threw out food, bought clothes that were never worn, missed a train or bus, had a mortgage foreclosed, had a driver's license revoked, had an unplanned pregnancy) and inversely weighted by their frequency.

PREJUDICES ABOUT BIASES—AND THE RHETORIC OF COMPETENCE

Over the past 40 years, the study of judgment and decision making has spread widely, first to social psychology,[70] then to application areas like accounting, health, and finance, finally penetrating mainstream economics under the banner of behavioral economics. That success owes something to the power of the approach, which liberated researchers previously bound by rational-actor models for describing

TABLE 2 | Correlations Between Decision-making Competence (DMC) and Other Variables

		Semi-partial correlation, controlling for		
DMC correlated with	Pearson *r*	Vocabulary	ECF	Vocabulary and ECF
Cognitive ability				
Vocabulary	.50	—	.28	—
ECF	.48	.26	—	—
Overall*	p < .0001	p = .0009	p = .0008	—
Cognitive style				
Polarized thinking	−.34	−.20	−.24	−.19
Self-consciousness	.20	.14[b]	.05	.11
Self-monitoring	.24	.29[b]	.30[b]	.32
Behavioral coping	.32	.27[a]	.28[a]	.26
Overall*	p < .0001	p < .0001	p < .0001	p < .0001
Risk behavior				
Antisocial disorders	−.19	−.18[b]	−.05	−.09
Externalizing behavior	−.32	−.28[b]	−.18	−.20
Delinquency	−.29	−.28[b]	−.18	−.21
ln(lifetime # of drinks)	−.18	−.22[b]	−.15	−.18
ln(lifetime marijuana use)	−.25	−.30[b]	−.20	−.25
ln(# times had sex)	−.24	−.30[b]	−.21	−.27
ln(# sexual partners)	−.30	−.33[b]	−.29[a]	−.31
Overall*	p = .0004	p = .0002	p = .009	p = .002
Social and family influences				
Risk status (HAR = 1; LAR = 0)	−.35	−.27	−.23	−.21
SES	.35	.20	.21	.15
Social support	−.30	−.21	−.23	−.19
Positive peer environment	.33	.35[b]	.32[a]	.35
Overall*	p = .0002	p = .002	p = .006	p = .007

ECF = executive cognitive function; HAR = high risk family; LAR = low risk family; SES = socioeconomic status.
[a]Test A rejects the one-mediator null hypothesis.
[b]Test B rejects the one-mediator null hypothesis.
Source: Parker and Fischhoff (2005),[68] where the tasks are described more fully.

behavior. It also owes something to the fascination of results that address a central aspect of the human condition, individuals' competence to manage their own affairs.[50,66,67] Very different social institutions may suit rational actors (e.g., free markets, civic engagement) and irrational ones (e.g., strong regulatory protection, deference to paternalistic experts).

Those seeking to extract general messages from this complex research literature have adopted several archetypal rhetorical stances. Familiarity with these stances can help in seeing the research through the stances. Table 3 summarizes several common themes, formulated in terms of their advocates' interpretation of the demonstrations of bias that tend to dominate the field.

It Is Not True

Examining research for possible flaws is central to any science. However, as seen in Table 1, the set of features that might, conceivably change a research result is very large, allowing endless criticisms by those who dislike a result ('the bias would disappear had you just changed...'). Such radical skepticism may be met by radical counter-skepticism ('you can't test for every conceivable confound'). A compromise asks whether confounds have general effects. The 'unfair tasks' section of Table 3 lists common methodological criticisms (e.g., biases would vanish with higher stakes or clearer instructions). An early review of all studies studying these factors found no effect on hindsight bias or on overconfidence in beliefs. A more recent review found that financial incentives had mixed

TABLE 3 | Debiasing Methods According to Underlying Assumptions

Assumption	Strategies
Faculty tasks	
Unfair tasks	Raise stakes
	Clarify instructions
	Dispel doubts
	Use better response modes
	Discourage second guessing
	Ask fewer questions
Misunderstood tasks	Demonstrate alternative goal
	Demonstrate semantic disagreement
	Demonstrate impossibility of task
	Demonstrate overlooked distinction
Faulty judges	
Perfectible individuals	Warn of problems
	Describe problem
	Provide personalized feedback
	Train extensively
Incorrigible individuals	Replace them
	Recalibrate their responses
	Plan on error
Mismatch between judges and task	
Restructuring	Make knowledge explicit
	Search for discrepant information
	Decompose problem
	Consider alternative situations
	Offer alternative formulations
Reeducation	Rely on experts
	Educate from childhood

Source: Fischhoff (1982).[71]

effects, sometimes improving performance, sometimes degrading it, but most often making no difference.[13]

It Is True, But You Should Not Say So

Demonstrations of bias allow researchers, who claim to know the answers, to fault others, who do not. Charging others with incompetence undermines their right to make decisions. As a result, researchers should avoid sweeping statements about human competence—and stick to the details of domain-specific studies. They should convey both the 'figure' of biases and the 'background' of the heuristics producing them. They should recall that optical illusions reveal important properties of vision without hindering most activities.[21] They should resist those who promote their research because it serves their political ends.

People Are Doing Something Quite Different—And Doing It Quite Well

Describing decisions as suboptimal presumes a normative analysis, informed by knowledge of what people know and want. Without that analysis, evaluations can be unduly harsh (e.g., charging overconfidence when people have strategically overstated their beliefs) or lenient (e.g., excusing mistakes as attempts to learn by trial and error). Table 3 'misunderstood tasks' section lists some ways that actors and observers can interpret decisions differently. In experiments, *manipulation checks* can assess whether subjects understand tasks as intended. In the world, observers are typically left guessing. For example, there is an unresolved controversy over whether some Americans increased their travel risk, by driving rather than flying, right after the 9/11 attacks. However, the interpretation of their decisions requires knowing how they saw the costs, risks, and hassles of flying and driving. Without evidence on these beliefs, any evaluation of their choices is speculative.

But Look at How Well People Do Other Things

Claims of bias seem strikingly at odds with the complex tasks that people routinely accomplish (including driving and flying). Perhaps, the biases are just laboratory curiosities, theoretically informative, but of limited practical importance. Or, perhaps the research denies people supports that life typically affords them. Table 3 'restructuring tasks' section lists manipulations that have improved performance under lab conditions. For example, when prompted, people can generate reasons why they might be wrong (reducing overconfidence), ways that events might have turned out otherwise (reducing hindsight bias), or estimates of what normally happens (reducing base-rate neglect). If life provides similar cues, then these 'debiasing' studies are most relevant for extrapolation to actual behavior.

Facing the Problems

Arguably, by mid-adolescence, most people have the cognitive ability to acquire most of the skills needed to make better decisions.[52,67,69,72] Whether they do depend on the help that they get. Unfortunately, people often receive little training, feedback, and help in making decisions. Indeed, they often face marketers, politicians, and others trying to manipulate their choices.[44,73] Table 3 'perfectible individuals' section lists strategies that seem able to enhance individuals' decision-making abilities—recognizing that their success, in any specific setting, is an empirical question.[74,75] The 'incorrigible individuals' section lists ways to live with fallibility. A historical example of recalibration was doubling engineers' chronic underestimates of the repair time for power plants.[76] A currently popular compromise is 'nudging' people toward better decisions, by choosing better default choices (e.g., being an organ donor, contributing to pension plans).[50]

CONCLUSIONS

Judgment and decision making research both requires and allows an unusual degree of collaboration among scientists with diverse expertise. The core discipline of behavioral decision research entails familiarity with normative analyses, descriptive studies, and prescriptive interventions. Its execution involves input from experts in the subject matter of specific decisions, the other (social and affective) pressures on them, and the opportunities for change.[52] For example, Downs et al.[61] helped young women make better sex-related decisions, with an interactive DVD, whose content reflected medical research (about sexually transmitted infections), behavioral decision research (about risk perceptions), and social psychology (about self-efficacy).

Behavioral decision research also provides a research platform where theoretical and practical research is mutually reinforcing. In the study of clinical judgment, such interactions showed the predictive power of simple models, a result that was invisible to researchers immersed in domain-specific research. In the study of judgment under uncertainty, these interactions revealed suboptimal strategies that survive because they are good enough to avoid major problems. In the study of value elicitation, they revealed the constructive nature of preference formation, as individuals infer what they want in the novel situations created by life and researchers. In the study of choice, they revealed the positive and negative interplay of cognition and affect. The field's future may exemplify Allan Baddeley's[77] call for the integrated pursuit of *applied basic* research, testing theory by its application, and *basic applied* research, creating theory from new phenomena observed through those tests.

REFERENCES

1. Coombs CH, Dawes R, Tversky A. *Mathematical Psychology*. Englewood Cliffs, NJ: Prentice Hall; 1970.
2. Edwards W. A theory of decision making. *Psychol Bull* 1954, 54:380–397.
3. Berkeley D, Humphreys PC. Structuring decision problems and the "bias heuristic." *Acta Psychol* 1982, 5:201–252.
4. Hammond KR. *The Psychology of Egon Brunswik*. New York: Holt, Rinehart & Winston, 1966.
5. Fischhoff B, Beyth-Marom R. Hypothesis evaluation from a Bayesian perspective. *Psychol Rev* 1983, 90:239–260.
6. Goldberg LR. Simple models or simple processes? Some research on clinical judgments. *Am Psychol* 1968, 23:483–496.
7. Kelly EL, Fiske DW. *The Prediction of Performance in Clinical Psychology*. Ann Arbor: University of Michigan Press, 1951.
8. Meehl PE. *Clinical Versus Statistical Prediction: A Theoretical Analysis and A Review of the Evidence*. Minneapolis: University of Minnesota Press, 1954.
9. Dawes RM. The robust beauty of improper linear models in decision making. *Am Psychol* 1979, 34:571–582.
10. Dawes RM, Faust D, Meehl P. Clinical versus actuarial judgment. *Science* 1989, 243:1668–1674.
11. Ericsson A, Simon HA. *Verbal Reports as Data*. Cambridge, MA: MIT Press, 1994.
12. Nisbett RE, Wilson TD. Telling more than we know: Verbal reports on mental processes. *Psychol Rev* 1977, 84:231–259.
13. Camerer CF, Hogarth RM. The effects of financial incentives in experiments: A review and capital-labor-production framework. *J Risk Uncertainty* 1999, 19:7–42.
14. Becker MH. The• health belief model and personal health behavior. *Health Educ Monogr* 1974; 2(4). AQ2
15. Hochbaum G. Why people seek diagnostic x-rays. *Publ Health Rep* 1956, 71:377–380.
16. Ogden J. Some problems with social cognition models: A pragmatic and conceptual analysis. *Health Psychol* 2003, 22:424–428.
17. Fischhoff B. Cognitive processes in stated preference methods. In: Mäler K-G and Vincent J, eds. *Handbook of Environmental Economics*. Amsterdam: Elsevier; 2005, 937–968.

18. Viscusi WK. *Risk by Choice* Cambridge, MA: Harvard University Press, 1983.
19. Slovic P, Lichtenstein S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organ Behav Hum Perform* 1971, 6:649–744.
20. von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.
21. von Winterfeldt D, Edwards W. *Decision Analysis and Behavioral Research*. New York: Cambridge University Press, 1986.
22. Clemen RT. *Making Hard Decisions: An Introduction to Decision Analysis*. Belmont, CA: Duxbury, 2003.
23. Peterson CR, Beach LR. Man as an intuitive statistician. *Psychol Bull* 1967, 68:29–46.
24. Jonides J, Naveh-Benjamin M. Estimating frequency of occurrence. *J Exp Psychol Hum Learn Mem* 1987, 13(2):230–240.
25. Hasher L, Zacks RT. Automatic processing of fundamental information. *Am Psychol* 1981, 39:1372–1386.
26. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science* 1974, 185:1124–1131.
AQ3 27. Gilovich T, Griffin D, Kahneman D eds. *Heuristics and Biases:• The Psychology of Intuitive Judgment*. New York: Cambridge University Press; 2002.
AQ4 28. Kahneman D, Slovic P, Tversky A. *Judgment Under• Uncertainty: Heuristics and Biases*. Cambridge University Press; 1982.
29. Edwards W, Lindman HR, Savage LJ. Bayesian statistical inference for psychological research. *Psychol Rev* 1963, 70:193–242.
30. Phillips LD. *Bayesian Statistics for Social Science* London: Nelson, 1973.
31. Fischhoff B. Risk perception and communication. In: Detels R, Beaglehole R, Lansang MA, Gulliford M, eds. *Oxford Textbook of Public Health*. 5th ed. Oxford: Oxford University Press; 2009, 940–952.
32. Poulton EC. The new psychophysics: Six models for magnitude estimation. *Psychol Bull* 1968, 69:1–19.
33. Poulton EC. *Behavioral Decision Making* Hillsdale, NJ: Lawrence Erlbaum, 1994.
34. Stevens SS. *Psychophysics: Introduction to Its Perceptual, Neural and Social Prospects*. New York, NY: John Wiley & Sons, 1975.
35. Tune GS. Response preferences: A review of some relevant literature. *Psychol Bull* 1964, 61:286–302.
36. Lichtenstein S, Slovic P, Fischhoff B, Layman M, Combs B. Judged frequency of lethal events. *J Exp Psychol Hum Learn Mem* 1978, 4:551–578.
37. Ajzen I, Fishbein M. *Understanding Attitudes and Predicting Social Behavior* Englewood Cliffs, NJ: Prentice-Hall, 1980.
38. Gilbert DT. *Stumbling on Happiness*. New York: Knopf, 2006.
39. Fischhoff B. Value elicitation: Is there anything in there? *Am Psychol* 1991, 46:835–847.
40. Lichtenstein S, Slovic P, eds. *Construction• of Preferences*. New York: Cambridge University Press, 2006. AQ5
41. Lupia A. Shortcuts versus encyclopedias—information and voting behavior in California insurance reform elections. *Am Polit Sci Rev* 1994, 88:63–76.
42. Schläpfer F. Contingent valuation: a new perspective. *Ecol Econ* 2008, 64:729–740.
43. Politi MC, Han PKJ, Col N. Communicating the uncertainty of harms and benefits of medical procedures. *Med Decis Making* 2007, 27:681–695.
44. Schwartz LM, Woloshin S, Welch HCG. Using a drug facts box to communicate drug benefits and harms. *Annals of Internal Medicine* 2009, 150:516–527.
45. McGuire W. Suspiciousness of experimenter's intent. In: Rosenthal R, Rosnow RL, eds. *Artifact in Behavioral Research*. 1969, New York: Academic Press.
46. Fischhoff B, Slovic P, Lichtenstein S. Knowing what you want: Measuring labile values. In: Wallsten T, ed. *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, NJ: Erlbaum; 1980, 117–141.
47. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. *Econometrica* 1979, 47:263–281.
48. Helson H. *Adaptation Level Theory: An Experimental and Systematic Approach to Behavior* New York: Harper & Row, 1964.
49. Johnson EJ, Goldstein D. Do defaults save lives?. *Science* 2003, 302:1338–1339.
50. Thaler RH, Sunstein CR. *Nudge* New Haven: Yale University Press, 2008.
51. Fischhoff B. Predicting frames. *J Exp Psychol Learn Mem Cogn* 1983, 9:113–116.
52. Reyna VF, Farley F. Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. *Psychol Sci Publ Interest* 2006, 7:1–44.
53. Simon HA. *Models of Man*. New York: John Wiley & Sons, 1957.
54. Gigerenzer G, Selten R. *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press, 2001.
55. Johnson EJ, Haubl G, Keinan A. Aspects of endowment: A Query Theory of value construction. *J Exp Psychol* 2007, 33:461–474.
56. Tversky A, Koehler DJ. Support theory: A nonextensional representation of subjective probability. *Psychol Rev* 1994, 101:547–567.
57. Weber EU, Johnson EJ. Mindful• judgment and decision making. *Annu Rev Psychol*; 60:53–85. AQ6
58. Lerner JS, Keltner D. Beyond valence: Toward a model of emotion-specific influences on judgment and choice. *Cognit Emot* 2000, 14:473–493.

59. Lerner JS, Small DA, Fischhoff B. Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychol Sci* 2003, 14:144–150.
60. Loewenstein G, Weber E, Hsee C, Welch N. Risk as feelings. *Psychol Bull* 2001, 67:267–286.
61. Slovic P, Peters E, Finucane ML, MacGregor D. Affect, risk and decision making. *Health Psychol* 2005, 24:S35–S40.
62. Tribe L. Ways not to think about plastic trees. *Yale Law J* 1974, 83:1315–1346.
63. Downs JS, Murray PJ, Bruine de Bruin W, White JP, Palmgren C. et al. An interactive video program to reduce adolescent females' STD risk: A randomized controlled trial. *Soc Sci Med* 2004, 59:1561–1572.
64. Thorngate W. Efficient decision heuristics. *Behav Sci* 1980, 25:219–225.
65. Ben-Haim, Y. *Info-gap Decision Theory: Decisions Under Severe Uncertainty*. 2nd ed. London: Academic Press; 2006.
66. Todd PM, Gigerenzer G. *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press, 2009.
67. Stanovich KE. *Decision Making and Rationality in the Modern World* New York: Oxford University Press, 2009.
68. Parker A, Fischhoff B. Decision-making competence: External validity through an individual-differences approach. *J Behav Decis Making* 2005, 18:1–27.
69. Bruine de Bruin W, Parker A, Fischhoff B. Individual differences in adult decision-making competence (A-DMC). *J Pers Soc Psychol* 2007, 92:938–956.
70. Nisbett RE, Ross L. *Human Inference: Strategies and Shortcomings of Social Judgment* Englewood cliffs, NJ: Prentice-Hall, 1980.
71. Fischhoff B. Debiasing. In: Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press; 1982, 422–444.
72. Fischhoff B. Assessing adolescent decision-making competence. *Dev Rev* 2008, 28:12–28.
73. Schwartz LM, Woloshin S, Welch HCG. *Know Your Chances* Berkeley, CA: University of California Press, 2008.
74. Baron J, Brown R, eds. *Teaching• Decision Making to Adolescents*. New Jersey: Erlbaum; 1990, 19–60. AQ7
75. Larrick RP, Morgan JN, Nisbett RE. Teaching the use of cost-benefit reasoning in everyday life. *Psychol Sci* 1990, 1:362–370.
76. Kidd JB. The utilization of subjective probabilities in production planning. *Acta Psychol* 1970, 34:338–347.
77. Baddeley AD. Applied• cognitive and cognitive applied research. In: Nilsson LG, ed. *Perspectives on Memory Research*. Hillsdale, NJ: Lawrence Erlbaum; 1979. AQ8

FURTHER READING

Fischhoff B, Kadvany J. *Risk: A Very Short Introduction*. London: Oxford University Press; 2010.
Gilovich T, Griffin D, Kahneman D, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press; 2002.
Hastie R, Dawes RM. *Rational Choice in an Uncertain World*. San Diego: Russell Sage; 2002.
Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press; 1982.
Kahneman D, Tversky A, eds. *Choices, Values, and Frames*. New York: Cambridge University Press; 2000.
Lichtenstein S, Slovic P, eds. *Construction of Preferences*. New York: Cambridge University Press; 2006.
von Winterfeldt D, Edwards W. *Decision Analysis and Behavioral Research*. New York: Cambridge University Press; 1986.
Yates JF. *Judgment and Decision Making*. New York: John Wiley & Sons; 1990.

