CHAPTER 3

Risk-Based Thinking

Baruch Fischhoff

For 15 years, I headed an undergraduate major in Decision Science. It has been successful enough that, in the College of Humanities and Social Sciences, it lags only economics and psychology in popularity among disciplinary majors. About one fifth of entering students list it as a possible major—a startling percentage given that no high school teaches decision science. It also gets many "refugees" from other departments (e.g., physics, engineering, architecture), who find that Decision Science provides a better balance of quantitative and qualitative approaches to basic and applied problems. Some students find the major satisfying enough to stay in touch after graduation. One, when visiting campus during Spring Carnival, observed that "I can't turn it off." He went on to describe instances in which he saw the world differently due to his education.

It was a gratifying comment for an educator. Perhaps we, collectively, have been doing our part in meeting the demand attributed to H. G. Wells: "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."[1] However, it was also a sobering comment, raising the empirical question of how our curriculum had, in fact, affected our students' decision-making competence and confidence in those abilities.

The same question applies to our field overall. In 40-some years, it has grown dramatically from its origin as a handful of individuals committed to doing research that was both analytically and behaviorally informed, with the dual missions of contributing both to science and public welfare (Slovic, Kunreuther, and White 1974; Slovic, Fischhoff, and Lichtenstein 1976; Jungermann and DeZeeuw 1977; Kunreuther et al. 1978). "The List" maintained by

Sarah Lichtenstein had fewer than 100 members, including some behind the Iron Curtain whose work was restricted because they used terms like "subjective" and "utility." Today, the field is represented in societies, journals, meetings, agencies, consultancies, and regulations. It is disseminated through classes, texts, websites, TED talks, and trade books that include nonfiction bestsellers (Ariely 2008; Thaler and Sunstein 2008; Kahneman 2011). In thinking about the future of risk research, it is important to ask how well we have fulfilled our version of George Miller's (1969) admonition to "give psychology away," for the common good.

Ideally, encounters with our work will make people better decision-makers, by helping them to recognize decision points, devote proper resources to them, estimate risks and benefits, balance intuitive and reflective responses, assess uncertainties in their beliefs, resolve ambiguities in their preferences, and defend themselves from needless regret, knowing that they have done what they could to make the best choices possible in complex, uncertain, and sometimes unfriendly circumstances.

Less ideally, encounters with our work will leave people worse off, by undermining their intuitive ways of thinking, drowning them in bewildering arrays of potential biases, without providing useful alternatives ways to deal with the risk decisions in their lives. Members of our public might end up unduly humbled by a "gotcha brigade" of researchers dedicated to highlighting human failings. Or, they might end up overly confident, convinced that they are now immune to the follies illustrated in others' behavior.

At the professional level, progress in risk management research and practice is obvious—witness the celebration of the Wharton School's Risk Management and Decision Processes Center that prompted this book. But what about the spillover effects to everyday life? To what extent do people, as individuals and society, think more clearly about risks, due to our work? The answer to that question should set our agenda for creating and applying the needed science. Those applications might include providing people with better information, helping them to articulate their preferences, clarifying the positive and negative roles of emotion, and protecting them from deceptive advertising and unsafe products. When risk professionals manage to provide that support, they put themselves out of business, having allowed people to fend for themselves. They can then move on to address other, more difficult decisions.

The next two sections discuss the challenges of assessing how well people think about risks. The first focuses on the question of how well people make

specific choices, the second on the question of how well they have mastered the skills needed for decision-making in general. The following section considers ways to enhance the transfer of knowledge from the professional world of risk research to the practical world in which people make risk decisions and live with their consequences. Each section leads to a proposal for a strategic research initiative. The concluding section asks how the field can organize itself to address these opportunities.

How Well Can People Think about Specific Risk Decisions?

The title of Dan Ariely's well-known *Predictably Irrational* (2008) captures one general pattern found in risk research: people are, to some extent, predictably irrational, in the sense that they make judgments and decisions contrary to normative accounts of how they should behave. The title of Daniel Kahneman's *Thinking, Fast and Slow* (2011) captures a fundamental distinction that can guide predictions of such irrationality: knowing whether people are thinking fast or slow. Fast thinking can produce more irrational decisions (in Ariely's sense) when people who stop to think have better heuristics in their cognitive repertoire than the ones that immediately come to mind. Slow responses can produce more irrational decisions when people have learned, or instinctively know, what to do, so that reflection wastes time or leads them astray. The better our science, the better we can predict the roles and outcomes of fast and slow thinking. The title of Paul Slovic's important collection *The Feeling of Risk* (2001) raises analogous questions about how the feeling of risk can direct and misdirect decisions.

Although clearly conditional, these accounts are often cited as making universal statements about how well people think about risk and uncertainty. That oversimplification changes "predictably" from a property of situations, some of which produce irrational responses, to a property of individuals, who bring irrationality wherever they go. Slow thinking is always better. Feeling is always problematic. Accentuating the negative in these ways reflects a figure-ground effect; problems (biases) are more salient than the processes that produce them (heuristics). Focusing on problems can serve the public if it directs researchers to places where help is needed. It can also serve researchers' egos, by making them the arbiters of others' limitations.

There are procedures for evaluating the optimality of any specific choice and, to a lesser extent, the rationality of the processes leading to it (von Winter-

feldt and Edwards 1986; Fischhoff and Eggers 2012; von Winterfeldt 2013). These procedures ask to what extent people manage to choose options in their own best interest, despite any limits to their knowledge, cognitive abilities, and affective control. The preferred option is typically defined as some variant of that "having the greatest expected utility." Performing that calculation requires understanding how individuals formulate their choices; for example, it should capture cases where people are willing to bear the consequences of poor choices in return for the benefits of learning from experience (Einhorn 1986; Baron 1994). The best option could be one that violates the expected utility standard, if it enables decision-makers to defend a "sacred" (or "protected") value, of the sort that allows no tradeoffs (Baron and Spranca 1997; Atran and Axelrod 2008). It could also be one that reflects nonconsequentialist goals, such as wanting to choose independently (e.g., for adolescents or the elderly) or to be ruled by passion (e.g., for love or war).

Thus, although conceptually straightforward, assessing the quality of individuals' decision-making is technically demanding. It requires taking the broad view needed to identify the kinds of decisions that individuals want to make. There are many thoughtful reflections on "what makes a good decision" that could inform that assessment (Yates, Veinott, and Patalano 2003; Baron 2008). However, there is no systematic procedure for applying them, sensitive to the potential diversity in individuals' preferences for decision-making consequences and processes. In its absence, observers risk incomplete accounts. They may rush to judgment, wrongly concluding that decision-makers cannot understand risks, leading to policies that manipulate people who could manage their own affairs. Conversely, observers may be overly creative in justifying decision-makers' choices and competence, leading to policies that deny them needed protection in situations where they cannot fend for themselves.

Risk Research Need 1: Procedures for Evaluating How Well People Make Risk Decisions

Various authors have offered frameworks accounting for subsets of the factors potentially shaping risk decisions. For example, Grether and Plott (1979) proposed a taxonomy of potential artifacts in experimental choice tasks. Lerner and Tetlock (1999) provided one for factors determining how accountable people feel for their choices. Milkman, Chugh, and Bazerman (2009) have a

related proposal for characterizing debiasing procedures. Lita Furby and I created a framework for specifying the prospects offered in stated preference studies (Fischhoff and Furby 1988; Fischhoff 2005). Florig et al. (2001) developed a procedure that characterizes diverse risks in common terms. When applied consistently, such schemes fulfill some of Gary Becker's (1976) call to ensure that preferences are stable across related decisions, an assumption underlying revealed preference analyses. Such frameworks reflect past research by focusing on factors that have been found to affect choices. They protect future research by circumscribing the set of potential explanations, thereby reducing the risk of ad hoc interpretations. By putting diverse studies on a common footing, such frameworks structure the kind of rolling meta-analyses that can provide the transparency advocated by the open science movement (Braver, Thoemmes, and Rosenthal 2014). Because risk management addresses real problems, the set of potentially relevant factors is naturally diverse. Having shared procedures for evaluating decisions would provide a foundation for orderly interventions and accumulation of knowledge.

How Can We Tell How Well People Think about Risk Decisions Generally?

Even were there authoritative accounts of how well people make specific risk decisions, that alone would not answer the question of how well people make risk decisions overall—and whether they are getting better over time. Such a general assessment could mean attempting to weight those specific decisions in terms of which matter most. Is it the repeated ones or the unique ones? The easy ones or the hard ones? Decisions involving heart and soul, or pocketbook? Researchers in healthcare services must make such cross-decision comparisons when allocating societal resources. Others need not.

An alternative question asks to what extent individuals have mastered the general skills needed to make risk decisions. If those skills have increased over time, then risk research might take some credit. The "Flynn effect" (Flynn 1987, 2009) provides reason for optimism regarding the skills defined by intelligence tests. Those scores appear to be going up over time. The vigorous debate over the meaning of these changes captures the challenges of such assessments. The internal validity of widely used IQ tests has been studied extensively. However, their external validity remains open to question. To what

extent do they assess the skills that people need to succeed in life, rather than just test-taking abilities that are correlated with success? Do the tests create an illusion of validity, or self-fulfilling prophecy, by directing resources to people who score well, thereby helping them to succeed (Einhorn 1982)?

Tests of individual differences in decision-making competence are a long way from affecting college admissions, the way that IQ tests do. Indeed, until recently, they hardly existed at all. There have long been tests of individual differences in cognitive style. However, those had such limited predictive validity that George Huber (1983) once proposed abandoning the search, arguing that anything big enough to have practical value would have been found already, then followed his own advice by moving on to study organizational change.

One source of researchers' disinterest in individual differences was Walter Mischel's (1968) account of how situational factors often overwhelm personality factors. Psychologists codified the tendency to neglect situational factors when interpreting others' behavior as the "fundamental attribution error" (Ross 1977). Another source of researchers' disinterest was their focus on widely shared psychological processes, relevant to creating a general picture of how people think. A third source was experimentalists' need to vary tasks across studies, when probing the effects of situational factors, rather than standardizing their tasks, as required for individual difference measures. A fourth source was having different participants in most studies, which avoids learning effects, but forfeits the chance to observe behavior over time and tasks.

In our own research, an opportunity arose to add a battery of common decision-making tasks to a 20-year longitudinal project, the Center for Education and Drug Abuse Research, led by Ralph Tarter. We found that performance on these tasks was correlated, suggesting a common factor of decision-making competence (DMC). Moreover, DMC scores were correlated with measures of plausible antecedents and consequences, suggesting the external validity of our tasks. For example, participants with higher DMC scores were more likely to come from intact homes and less likely to exhibit oppositional defiance disorder and other risk behaviors. Moreover, those correlations remained after controlling for scores on tests of fluid and crystalized intelligence (Parker and Fischhoff 2005). These initial patterns have generally borne up in subsequent research (Bruine de Bruin, Parker, and Fischhoff 2007b; Parker, Bruine de Bruin, and Fischhoff 2007, 2015; Missier et al. 2015), including evidence of stability over an 11-year period (Parker et al.

2018). Similar patterns have also emerged in concurrent research into the skills demanded by reasoning tasks and imperfectly captured by intelligence tests (Stanovich and West 2000, 2008).

One sustained attempt to track the quality of lay thinking over time is the National Science Foundation's survey of "science literacy" (National Science Board 2014). Although motivated by the desire to make science more useful, these surveys have not attempted to assess decision-making abilities or scientific reasoning skills (Drummond and Fischhoff 2015). Rather, they have evaluated respondents' knowledge of facts that might be interpreted as markers of those abilities (e.g., "The center of the earth is hot"). As a result, they say relatively little about the extent to which the seeming rise in public skepticism about science reflects less ability to reason like scientists or less willingness to believe the factual premises underlying scientists' reasoning (Lewandowsky et al. 2012).

Research Need 2: A Shared Set of Validated Decision-Making Competence Measures

Assessing changes in decision-making skills over time requires standard instruments and places to deploy them. In creating our measure of DMC (Parker and Fischhoff 2005; Bruine de Bruin et al. 2007b), we included tasks that would interest researchers concerned with specific tasks (e.g., calibration, framing) and the overall skill set. Although we modeled our tasks on ones in the literature, no scientist should presume to prescribe others' measures. One institutional model for balancing scientific freedom and standardization is the NIH-sponsored PROMIS (Patient-Reported Outcomes Measurement Information System) consortium. Its website hosts any measure that meets its criteria for empirically demonstrated psychometric validity.[2] As a bonus, it offers users support in the form of applying item response theory to produce compact question sets. Researchers who use PROMIS measures produce results that are readily comparable to those from other studies. If risk researchers developed a PROMIS-like set of canonical DMC measures, they would be positioned for the kind of big science needed to study individual differences properly. That means having a data collection operation that is large enough to examine patterns in cross-sectional analyses and stable enough to capture developments in individuals and cohorts over time.

Although risk researchers have helped to refine such studies (Bruine de Bruin, Parker, and Fischhoff 2007a; Bruine de Bruin et al. 2011; Bruine de Bruin and Fischhoff 2017), they have rarely been central to their creation (Mellers et al. 2015).

How Can We Communicate Our Work Best?

If people are making better decisions and revealing greater decision-making competence, that could reflect the cumulative impact of our research, our books, talks, and courses. More direct evidence is found in experimental tests of interventions intended to reduce judgmental biases. Such debiasing studies have had mixed results. Just hearing about biases has no apparent effect on performance. Some success has been found with training that involves prompt, unambiguous feedback, supplemented by explication of unintuitive processes (Milkman, Chugh, and Bazerman 2009; Morgan 2014; Mellers et al. 2015; Morewedge et al. 2015). Studies of how well and how long such training transfers to actual decision-making are limited—and difficult to conduct. Thus, we know little about how exposure to our research affects individuals' confidence in their abilities. To that end, our goal might be inspiring humility: speaking with enough pride to draw attention to our research, while creating realistic expectations for how much difference it makes, as befits the difficulty of the decisions that people face, the limits to our science, and the apprenticeship needed to master it.

The science of science communication has identified challenges common to conveying any science (Fischhoff and Scheufele 2013, 2014): audiences might not understand the methods of a specific science or of science itself. They might lack the substantive knowledge (Bruine de Bruin and Bostrom 2013) needed to make sense of new findings. They might not know how strong a field is overall or how far specific findings can be extrapolated. They might struggle to decipher scientists' disagreements and expressions of uncertainty, especially when the science is contested (e.g., vaccines, climate change). They might be so committed to their beliefs that they quickly explain away contrary results and embrace supporting ones (Lewandowsky et al. 2009; Corner, Whitmarsh, and Xenias 2012).

Outside of the classroom, scientists have few opportunities to educate their audiences about the basics of their science. Rather, they must take people

as they are, with diverse, imperfect backgrounds, and provide enough context for recipients to interpret science appropriately. Complaints about scientists who seem to contradict one another or hype their work suggests that such context is often missing (e.g., Kolata 2016; McKay 2016). One way to provide it is with standard reporting formats that present information in a predictable format that gradually educates audiences in how scientists think about their work. The following exhibits offer four such approaches. As with all communications, their usefulness is an empirical question, answered more easily for immediate impacts (do users understand the content of the communication? can they use it to make sound inferences?) than for long-term ones (are they wiser decision-makers?).

General Disclosure

In *Risk: A Very Short Introduction* (Fischhoff and Kadvany 2011), we briefly reported results from many behavioral experiments, within a page limit that precluded elaborating on the strengths and weaknesses of each. Rather, the book offered general guidance, in the list that follows, on how the conditions of an experiment could affect the quality of the performance that it reveals. Each factor is common knowledge for scientists, but perhaps not to consumers of their work. Given the diverse studies that we reported, we made no attempt to assess the overall magnitude, or even sign, of these effects. However, researchers could do that for their own individual studies, providing readers with guidance on whether they observed particularly good or bad performance.

1. [Research tasks] are clearly described, so that researchers can see how people make them. That clarity can produce better decisions, if it removes the clutter of everyday life, or worse decisions, if that clutter provides vital context, such as what choices other people are making.
2. [Research tasks] have low stakes, reflecting researchers' limited budgets. That can produce better decisions, if it reduces stress, or worse decisions, if it reduces motivation.
3. [Research tasks] are approved by university ethics committees. That can produce better decisions, if it reduces participants' worry about being deceived, or worse decisions, if it induces artificiality.

4 [Research tasks] are focused on researchers' interests. That can produce worse decisions, if researchers are studying biases, or better decisions, if researchers are seeking decision-makers' secrets of success. (Fischhoff and Kadvany 2011, 110)

Methodological Audit

As part of a project investigating behavioral responses to smart grid electricity technology, Davis et al. (2013) reviewed 32 field trials of interventions (e.g., in-home displays of electricity usage). Using the CONSORT protocol for Cochrane Collaboration reviews, they characterized each study in terms of six methodological flaws found to affect medical clinical trials (Moher et al. 2010). For example, as seen in Figure 3.1, only four studies had low risk of volunteer selection bias, which can overestimate the impact of interventions—by studying people predisposed to change. The clinical trial literature has estimates of the size of the bias from some of these sources. Risk researchers could routinely characterize their research in such terms, winning points for their candor, while gradually educating their audiences about how science works. Fischhoff and Davis (2014) offer a detailed version of this audit, expanded to include the underlying the strength (or *pedigree*) of the underlying science (Funtowicz and Ravetz 1990).

Running Summary

No study stands alone. Its interpretation depends on knowledge of boundary conditions (list above) and methodological standards (Figure 3.1). When studies measure the same variable, meta-analysis can provide aggregate estimates of key variables (e.g., survival rates, sleep duration) (Braver, Thoemmes, and Rosenthal 2014). When studies measure related phenomena, funnel diagrams, like Figure 3.2, preserve the identity of the individual studies, while suggesting general patterns. In this example, the open triangles are from new studies. They reveal no overall effect of the experimental manipulation (x axis), and cluster closer to zero as the standard error decreases (y axis), as would be expected with orderly data. The gray cone shows the region for statistically nonsignificant results. The black circles are results

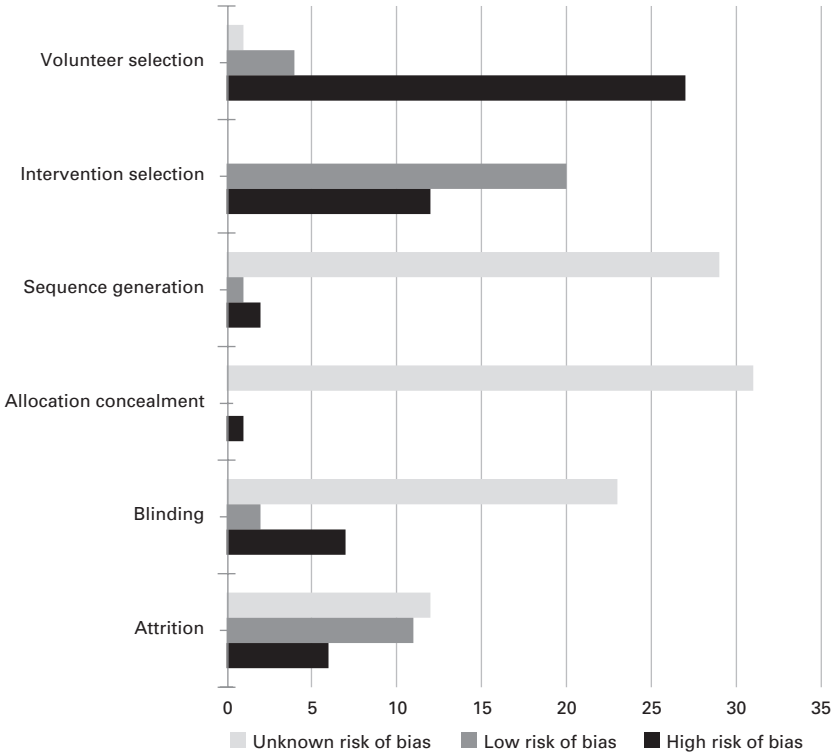

Figure 3.1. Methodological audit of field studies of interventions for reducing home electricity consumption. *Source:* Davis et al. (2013).

from 43 previous studies. They show an overall effect, but a disorderly one (outside the cone), leading the authors to question that research (Shanks et al., 2015). Although Figure 3.2 may be a daunting display at first glance, its elements are conceptually simple (e.g., how precise are the estimates, how anomalous is the pattern), meaning that it should possible to explain it, with some design work.

Integrative Decision Frame

When studies inform theories, the quality of the work is paramount (list above and Figure 3.2). When studies inform decisions, their practical relevance matters as well. Figure 3.3 is a tabular format recently adopted by the

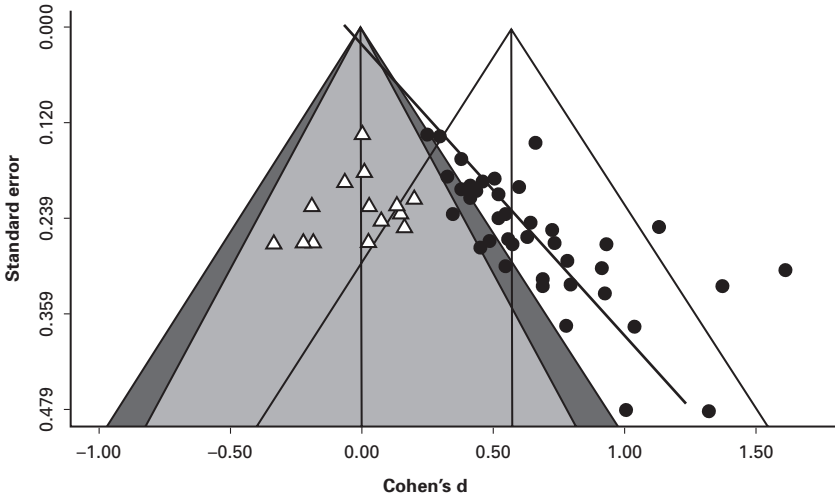

Figure 3.2. Funnel plots of related experimental studies. Cohen's d measures effect size for the difference between two groups (Cohen 1992). *Source:* Shanks et al. (2015).

Decision Factor	Evidence and Uncertainties	Conclusions and Reasons
Analysis of condition		
Current treatment options		
Benefit		
Risk		
Risk management		

Benefit-Risk Summary Assessment

Figure 3.3. The U.S. Food and Drug Administration's benefit-risk framework. *Source:* FDA (2013).

U.S. Food and Drug Administration (FDA) to summarize its evaluations of pharmaceuticals submitted for approval (Fischhoff 2017). The rows show how FDA frames its decisions. Namely, its risk-benefit decisions depend on the (medical) condition, the current treatment options, the expected benefits, the attendant risks, and the opportunities for managing risks (post-approval). The columns show that FDA distinguishes questions of science (on the left) from their regulatory implications (on the right). The former explicitly

recognize uncertainty (Institute of Medicine 2014). The latter depend on FDA's legal framework and understanding of its public (see also Schwartz and Woloshin 2013).

Research Need Number 3: Methods for Communicating the Quality of Risk Research

Researchers have tacit knowledge of how to interpret their work that observers, even those trained in other scientific disciplines, lack. Unless they provide that context, researchers risk evoking too much or too little credence for their studies. Like many decision science formulations, the four exhibits represent ways of organizing information that might not occur to people spontaneously but should be comprehensible if executed well. Recognizing that context should be good for the science as well. For example, Figure 3.2 gives greater credence to studies with orderly patterns of results, relative to studies with novel findings of uncertain provenance. The list above suggests how to make the case for "orchids," results that appear under highly specific conditions, hence might reveal interesting processes, even if they do not support grand claims about the human condition. Thus, developing standard ways to place risk research in context could improve the research itself, along with its contribution to aiding specific decisions and improving general decision-making competence.

Conclusion

The ultimate payoff for society's investment in risk research is better risk decisions. That goal can be advanced by addressing three research needs: (a) developing better procedures for evaluating the quality of individual decisions; (b) understanding decision-making competence, as it varies across individuals and over time; and (c) placing the science in context, so that observers know how to use it in their decision-making. The three research programs outlined here would advance these goals, contributing to the science and the society that supports it.

Our goal when interacting with any public might be the mature enthusiasm that we seek with our students: excitement about the research, tempered by recognition of its limits; wariness about biases, balanced by respect for the

power of heuristics; enthusiasm about the clear signals that experiments can offer, muted by the difficulty of generalizing their results; pride in the tension between simplicity and precision in theoretical accounts; the ability to think analytically, whatever one's level of technical mastery (National Research Council 2011).

People who can absorb, and live with, these tensions when consuming our work should come away with appropriate feelings of self-efficacy when making risk decisions. That perspective should afford them more satisfying lives, despite inevitable setbacks. It might give them an edge, in competition with over—or under—confident others. Another former student once returned a book with the comment, "I didn't realize that people thought that deeply." He was referring to a philosopher of science (Lakatos 1970). What if the consumers of our work thought the same about us?

To that end, the Wharton Risk Management and Decision Processes Center has seeded the world with people who speak the language of risk that it has helped to create. As the number of those people grows, they should increasingly find common cause, able to see and sustain solutions requiring their shared perspective. Over time, such thinking might become so natural that turning it off no longer is an option.

Notes

1. The less pithy actual quote was something like, "The time may not be very remote when it will be understood that for complete initiation as an efficient citizen of one of the new great complex world-wide states that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and to write" (Wells 1911, 204; see also Tankard 1979).

2. See HealthMeasures, "PROMRIS," www.healthmeasures.net/explore-measurement-systems/promis.

References

Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. New York: HarperCollins.

Atran, S., & Axelrod, R. (2008). Reframing sacred values. *Negotiation Journal, 24*, 221–246.

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences, 17*(1), 1–10.

Baron, J. (2008). *Thinking and deciding*. New York: Cambridge University Press.

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70*, 1–16.

Becker, G. (1976). Introduction. In *The economic approach to human behavior* (pp. 3–14). Chicago: University of Chicago Press.

Braver, S., Thoemmes, F., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science, 9*(3), 333–342.

Bruine de Bruin, W., & Bostrom, A. (2013). Assessing what to address in science communication. *Proceedings of the National Academy of Sciences, 110* (Supplement 3), 14062–14068.

Bruine de Bruin, W., & Fischhoff, B. (2017). Eliciting probabilistic expectations: Collaborations between psychologists and economists. *Proceedings of the National Academy of Sciences, 114*(13), 3297–3304.

Bruine de Bruin, W., Manski, C. F., Topa, G. & van der Klaaw, W. (2011). Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics, 26*, 454–478.

Bruine de Bruin, W., Parker, A., & Fischhoff, B. (2007a). Can adolescents predict significant events in their lives? *Journal of Adolescent Health, 41*, 208–210.

Bruine de Bruin, W., Parker, A., & Fischhoff, B. (2007b). Individual differences in adult decision-making competence (A-DMC). *Journal of Personality and Social Psychology, 92*, 938–956.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.

Corner, A., Whitmarsh, L., & Xenias, D. (2012). Uncertainty, scepticism, and attitudes towards climate change: Biased assimilation and attitude polarisation. *Climate Change, 114*, 463–478.

Davis, A. L., Krishnamurti, T., Fischhoff, B., & Bruine de Bruin, W. (2013). Setting a standard for electricity pilot studies. *Energy Policy, 62*, 401–409.

Drummond, C., & Fischhoff, B. (2015). Development and validation of the scientific reasoning scale. *Journal of Behavioral Decision Making, 30*(1), 26–38.

Einhorn, H. J. (1982). Learning from experience and suboptimal rules in decision making. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 268–283). New York: Cambridge University Press.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50*(3), 387–95.

FDA. (2013). *Structured approach to benefit-risk assessment for drug regulatory decision making*. Draft PDUFA V implementation plan, FY2013–2017. Silver Spring, MD: U.S. Food and Drug Administration.

Fischhoff, B. (1992). Giving advice: Decision theory perspectives on sexual assault. *American Psychologist, 47*, 577–588.

Fischhoff, B. (2005). Cognitive processes in stated preference methods. In K.-G. Mäler & J. Vincent (Eds.), *Handbook of environmental economics* (pp. 937–968). Amsterdam: Elsevier.

Fischhoff, B. (2013). The sciences of science communication. *Proceedings of the National Academy of Sciences, 110* (Supplement 3), 14033–14039.

Fischhoff, B. (2017). Breaking ground for psychological science: The U.S. Food and Drug Administration. *American Psychologist, 72*(2), 118–125.

Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences, 111* (Supplement 4), 13664–13671.

Fischhoff, B., & Eggers, S. (2012). Questions of competence: The duty to inform and the limits to choice. In E. Shafir (Ed.), *The behavioral foundations of policy* (pp. 217–230). Princeton: Princeton University Press.

Fischhoff, B., & Furby, L. (1988). Measuring values: A conceptual framework for interpreting transactions. *Journal of Risk and Uncertainty, 1*, 147–184.

Fischhoff, B., & Kadvany, J. (2011). *Risk: A very short introduction*. Oxford: Oxford University Press.

Fischhoff, B., & Scheufele, D. (Eds.). (2013). The science of science communication. *Proceedings of the National Academy of Sciences, 110* (Supplement 3).

Fischhoff, B., & Scheufele, D. (Eds.) (2014). The science of science communication II. *Proceedings of the National Academy of Sciences, 111* (Supplement 4).

Florig, H. K., Morgan, M. G., Morgan, K. M., Jenni, K. E., Fischhoff, B., Fischbeck, P. S., & DeKay, M. (2001). A deliberative method for ranking risks (1): Overview and test bed development. *Risk Analysis, 21*, 913–922.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191.

Flynn, J. R. (2009). *What is intelligence: Beyond the Flynn effect.* Cambridge: Cambridge University Press.

Funtowicz S. O., & Ravetz, J. (1990). *Uncertainty and quality in science for policy*. London: Kluwer.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review, 69*(4), 623–638.

Huber, G. P. (1983). Cognitive style as a basis for MIS and DSS designs: Much ado about nothing? *Management Science, 29*, 567–577.

Institute of Medicine. (2014). *Uncertainty in assessing pharmaceutical risks and benefits*. Washington, DC: National Academy Press.

Jungermann, H., & deZeeuw, G. (Eds.). (1977) *Decision making and change in human affairs*. Dordrecht: Reidel.

Kahneman, Daniel. 2011. *Thinking, fast and slow.* New York: Farrar Straus & Giroux.

Kolata, G. (2016, August 11). We're so confused: The problems with food and exercise studies. *New York Times.* Retrieved from https://www.nytimes.com/2016/08/11/upshot/were-so-confused-the-problems-with-food-and-exercise-studies.html.

Kunreuther, H., Ginsberg, R., Miller, L., Sagi, P., Slovic, P., Borkan, B., & Katz, N. (1978). *Disaster insurance protection: Public policy lessons*. New York: Wiley.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos and A. Musgrave (Eds.), *Criticism and the growth of scientific knowledge* (pp. 91–196). Cambridge: Cambridge University Press.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*(2), 255–275.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*, 106–131.

McKay, P. (2016, August 13). According to a recent study. . . . *Pittsburgh Post-Gazette.* Retrieved from http://www.post-gazette.com/life/2016/08/06/Peter-McKay-According-to-a-recent-study/stories/201608060020.

Mellers, C., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., . . . Tetlock, P. (2015). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science, 10*(3), 267–281.

Milkman, K. L., Chugh, D., & Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science, 4*(4), 379–383.

Miller, G. (1969). Psychology as a means of promoting human welfare. *American Psychologist, 24*, 1063–1075.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Missier, F. D., Hansson, P., Parker, A. M., Bruin de Bruine, W., Nilsson, L.-G., & Mäntylä, T. (2015). Unraveling the aging skein: Disentangling sensory and cognitive predictors of age-related differences in decision making. *Journal of Behavioral Decision Making*. doi:10.1002/bdm.1926

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., . . . Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ* 340:c869.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences, 2*, 129–140.

Morgan, M. G. (2014). The use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences, 111*, 7176–7184.

National Research Council. (2011). *Intelligence analysis for tomorrow*. Washington, DC: National Academy Press.

National Science Board. (2014). *Science and engineering indicators 2014* (NSB 14-01). Arlington, VA: National Science Foundation.

Parker, A. M., Bruine de Bruin, W., & Fischhoff. B. (2007). Maximizers vs. satisficers: Decision-making styles, competence and outcomes. *Judgment and Decision Making, 2*(6), 342–350.

Parker, A. M., Bruine de Bruin, W., & Fischhoff, B. (2015). Negative decision outcomes are more common among people with lower decision-making competence: An item-level analysis of the Decision Outcome Inventory (DOI). *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.00363.

Parker, A. M., Bruine de Bruin, W., Fischhoff, B., & Weller, J. (2018). Robustness of decision-making competence: Evidence from two measures and an 11-year longitudinal study. *Journal of Behavioral Decision Making, 31*, 309–470.

Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validity through an individual-differences approach. *Journal of Behavioral Decision Making, 18*, 1–27.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology, 10* (pp. 173–220). New York: Academic Press.

Schwartz, L., & Woloshin, S. (2013). The drug facts box: Improving the communication of prescription drug information. *Proceedings of the National Academy of Sciences, 110*, 14069–14074

Shanks, D. R., Vadillo, M. A., Riedel, B., Clymo, A., Govind, S., Hickin, N., . . . Puhlmann, L. M. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives? *Journal of Experimental Psychology: General, 144*(6), 142–158.

Slovic, P. (Ed.) (2001). *The feeling of risk*. London: Earthscan.

Slovic, P., Fischhoff, B. & Lichtenstein, S. (1976). Cognitive processes and societal risk taking. In J. S. Carroll and J. W. Payne (Eds.), *Cognition and social behavior* (pp. 165–184). Potomac, MD: Erlbaum.

Slovic, P., Kunreuther, H., & White, G. (1974). Decision processes, rationality and adjustment to natural hazards. In G. F. White (Ed.), *Natural hazards: Local, national and global*. New York: Oxford University Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Science, 23*, 645–726.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive abilities. *Journal of Personality and Social Psychology, 94*, 672–695.

Tankard, J. W. (1979). The H.G. Wells quote on statistics: A question of accuracy. *Historia Mathematica, 6*, 30–33.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New Haven, CT: Yale University Press.

von Winterfeldt, D. (2013). Bridging the gap between science and decision making. *Proceedings of the National Academy of Sciences, 110,* 14055–14061.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research.* New York: Cambridge University Press.

Wells, H. G. (1911). *Mankind in the making*, 5th ed. London: Chapman and Hall.

Yates, J. F., Veinott, E. S., & Patalano, A. L. (2003). Hard decisions, bad decisions: On decision quality and decision aiding. In S. L. Schneider & J. C. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 13–63). New York: Cambridge University Press.