26 Skilled Decision Theory: From Intelligence to Numeracy and Expertise

Edward T. Cokely
Department of Psychology, University of Oklahoma, Oklahoma

Adam Feltz
Cognitive and Learning Sciences, Michigan Technological University, Michigan

Saima Ghazal
Institute of Applied Psychology, University of the Punjab, Pakistan

Jinan N. Allan
National Institute for Risk & Resilience, and Department of Psychology, University of Oklahoma, Oklahoma

Dafina Petrova
Department of Experimental Psychology, Universidad de Granada, Spain

Rocio Garcia-Retamero
Department of Experimental Psychology, Universidad de Granada, Spain

Introduction

Over the past decade several landmark studies have advanced our scientific understanding of decision-making skill, its measurement, and its acquisition (i.e. *Skilled Decision Theory*). Here we present an integrative review of skilled human decision-making in experts and non- experts, with emphasis on four emerging insights.

(1) Among non-experts, normatively superior decision-making is associated with a domain-general skill that has largely been neglected in research on general intelligence. (2) Statistical numeracy tests (i.e. assessments of practical probabilistic reasoning) tend to be the strongest single predictors of *general decision-making skill* across wide-ranging numeric and non-numeric judgments and decisions (www.RiskLiteracy.org). (3) The superior decision-making exhibited by experts and non-experts primarily reflects specialized knowledge and integrated long-term memory representations that inform adaptive heuristic strategies (i.e. *representative understanding* rather than rational optimization). (4) High levels of basic cognitive abilities, such as fluid intelligence or attentional control, are not generally required for skilled or expert decision-making.

Although we've endeavored to minimize jargon in this chapter, some clarifications merit consideration. Historically, researchers have distinguished between *judgments* (e.g. estimates) and *decisions* (e.g. choices), based on traditions from the 1940s (e.g. decision researchers followed conventions in economics and statistics, while judgment researchers followed conventions in perception). Here, and for general purposes, the terms *judgment* and *decision-making* are roughly synonymous (e.g. a decision is a judgment about what to do; Baron, 2008). Thus, *general decision-making skill* refers to stable differences in judgment *and* decision-making quality exhibited across diverse and wide-ranging domains (e.g. health, wealth, and happiness). Likewise, verifiable *expert decision-making* refers to stable differences in judgment *and* decision-making quality exhibited within a specific domain (e.g. surgery, engineering, finance, chess). Following standard conventions, we use *general intelligence* as an umbrella term referring to a broad latent intelligence construct derived from interrelations among constituent basic cognitive abilities (e.g. fluid intelligence, crystallized intelligence, attentional control, memory, and others). We use *skill* to generally refer to acquired types of knowledge, skills, abilities, and related capacities. Finally, *bias* follows technical conventions referring to a tendency that does not necessarily imply error. Thus, biases may or may not be adaptive in various contexts (e.g. a "look left" bias when stepping into the street is much more adaptive in the United States than in the United Kingdom).

Our review includes five sections. First, we present a summary of the connections between statistical numeracy, general decision-making skill, and normative decision standards. Second, we review mechanisms that give rise to verifiable expert decision-making and skilled decision-making in general (Skilled Decision Theory). Third, we review psychometric studies of basic cognitive abilities, discussing when and why numeracy tests out-predict fluid intelligence tests. Fourth, we discuss simple decision aids and training programs that causally (and often dramatically) improve risk comprehension and skilled decision making. Finally, we close with a brief summary including ethical and policy implications.

Numeracy and Decision-Making Skill

Since the 1990s, research on the role of mathematical skills in decision-making has grown from practical efforts to improve risk communications and informed decision-making, particularly in health and medicine. Early work leveraged lessons from the National Assessment of Adult Literacy (Kutner, Greenberg, Jin, & Paulsen, 2006) and the Programme for International Student Assessment (Breakspear, 2012). A central focus was on the acquired skill *numeracy*, which refers to the "array of mathematically related proficiencies that are evident in adults' lives … including a connection to context, purpose, or use … for active participation in the democratic process and .. . in the global economy" (Ginsburg, Manly, & Schmitt, 2006). Accordingly, numeracy research is generally concerned with effective everyday problem-solving for activities like eval- uating medical treatment options, political claims, or financial products (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012; Cokely, Ghazal, Galesic, Garcia-Retamero, & Schulz, 2013; Cokely, Ghazal, & Garcia-Retamero, 2014; Newall, 2016; Reyna, Nelson, Han, & Dieckman, 2009; Steen, 1990). Based on theoretical and historical connections (Huff, 1954; Paulos, 1988), behavioral decision research on the role of numeracy began to rapidly advance following the introduction of a simple three-item psychometric numeracy test (Schwartz, Woloshin, Black, & Welch, 1997; see also Lipkus, Samsa, & Rimer, 2001). The seminal contributions by Schwartz and colleagues showed that (1) well-educated individuals often couldn't accurately answer basic numeracy questions (e.g. couldn't convert 1 in 1000 to 0.1 percent) and (2) numeracy scores robustly predicted the accuracy of disease risk interpretations. As of 2016, hundreds of studies have used variants of these classical-type numeracy tests to predict high-stakes real-world decisions.

Over the past five years, brief adaptive and item-response instruments have set the standard for numeracy assessment, including the most widely used modern tests validated for use with diverse samples from industrialized communities – i.e. the Berlin Numeracy Tests (Cokely et al., 2012, 2014; see also Garcia-Retamero & Cokely, 2017; Weller et al., 2013). Initial analyses based on 21 studies, including data from 15 countries, found that the Berlin Numeracy Tests were the strongest single predictors of individual differences in general decision-making skill, including the ability to evaluate and understand risk (i.e. risk literacy) across numerical and non-numerical evaluations of consumer products, medical treatments, and natural hazard forecasts. The three-minute adaptive Berlin Numeracy Test more than doubled the predictive power of the best available alternative numeracy instruments, uniquely predicting decision quality independent of general cognitive abilities (e.g. cognitive reflection, working memory, fluid intelligence). More than 100,000 people from 166 different countries have now taken one of the Berlin Numeracy Tests (adult skill sensitivity ranges from about the 10th to the 90th percentile in developed countries; see www.RiskLiteracy.org for tests and a test-version recommendation tool). Hundreds of subsequent studies with diverse participants from at least 60 countries – including surgeons, lawyers, patients, children, scientists, military veterans, police officers, athletes, older adults, nationally representative samples, and others – further document the unparalleled ability of *statistical numeracy* (i.e. practical probabilistic reasoning) to uniquely predict decision-making skill and risk literacy across (i) naturalistic, high-stakes, real-world choices (e.g. HIV prevention, cardiovascular risk mitigation, professional judgment of surgeons and physicians, public policy evaluations, natural hazard vulnerability, and many others) and (ii) theoretically essential paradigmatic tasks (e.g. risky prospect evaluation, framing resistance, sunk cost biases, recognizing norms, overconfidence, and others; Cokely et al., 2014; Garcia-Retamero & Cokely, 2011, 2013, 2014, 2017; Garcia-Retamero, Wicki, Cokely, & Hanson, 2014; Petrova, Kostopoulou, Delaney, Cokely, & Garcia-Retamero, 2017b).

The causal mechanisms linking numeracy and decision-making skill are manifold, including metacognitive, heuristic, intuitive, affective, subjective, gist-based, and number-sense processes (Cokely & Kelley, 2009; Ghazal, Cokely, & Garcia-Retamero, 2014; Lindskog, Winman, Juslin, & Poom, 2013; Peters et al., 2006; Peters & Bjalkebring, 2015; Peters, Hibbard, Slovic, & Dieckmann, 2007; Reyna, 2004, 2008; Schley & Peters, 2014; Thompson, Turner, & Pennycook, 2011; Traczyk & Fulawka, 2016). Accordingly, statistical numeracy is a robust predictor of numerical *and* non-numerical decisions, including judgments about social relationships, behavioral norms, professional competency, and many health behaviors (e.g. ignoring a heart attack). In part, this broad predictive power follows because statistical numeracy tests are themselves *representative judgment and decision-making tasks* that challenge inductive reasoning and self-regulation under conditions of risk and uncertainty. In other words, effective decision-making in our complex and uncertain world often requires the same kinds of reasoning and metacognitive skills that are used when solving various practical probabilistic math problems (e.g. evaluating thoughts, feelings, and risks). To further clarify connections, including shared cognitive and logical elements, we next consider standards of rationality and normative decision-making.

Rationality and Normative Standards

Decision science broadly involves three main projects: *descriptive* (e.g. what decisions do people make and why), *normative* (e.g. what decisions should be made and why), and *prescriptive* (e.g. how can actual decisions be improved and why) (Baron, 2008). Modern notions of *rationality* typically refer to *coherence standards* that may be used for the determination of normatively superior judgments and decisions. One of our most influential standards emerged in 1654 when Blaise Pascal began corresponding with Pierre de Fermat about the division of stakes in a popular gambling game. Although they were not the first to attempt to formalize *chance*, their letters became the founding documents of the logical system at the heart of modern science and decision theory – i.e. *probability theory* (Hacking, 2006). Exactly 300 years later a natural extension of inductive logic known as *Bayesian probability theory* enabled the axiomatization of *subjective expected utility theory* – i.e. the formal normative decision theory that is a basis of modern statistical and economic applications, as presented in the book *The Foundations of Statistics* (Savage, 1954; see also Jaynes, 2003; Schlaifer & Raiffa, 1961; Von Neumann & Morgenstern, 1944). To clarify, logic can be fundamentally divided into two major categories, namely (i) *deductive logic* – sound reasoning from premises to conclusions that are *certain* – and (ii) *inductive logic* – sound reasoning from premises to inferences that *involve risk* (e.g. characterized by known probability distributions) or *uncertainty* (e.g. characterized by unknown probabilities or exposure) (Holland, Holyoak, Nisbett, & Thagard, 1986; Savage, 1954). Because we live in a fundamentally risky and uncertain world, the practical decision-relevance of deductive logic typically pales in comparison to that of inductive logic.

Theoretically, the goal of all rational decision-making is to get more of what one should want (Baron, 1985, 2008; Hastie & Dawes, 2010). "Should" is complicated and requires many philosophical and value assumptions. Setting aside philosophical issues, good decisions can be defined by logical processes that coherently maximize desired outcomes in accord with integrative optimization techniques (e.g. determining the maximum or minimum value of a function subject to constraints as in formal cost–benefit or decision analysis). To illustrate, consider the example of deciding among several modest risky financial prospects such as lotteries involving two choices (e.g. (a) gain $100 for certain versus (b) 75 percent chance of gaining $200). Given enough choices like these and some other basic assumptions, on average a normatively superior decision would result when selecting options with the highest expected value – i.e. multiplying the probability of an occurrence by its value (e.g. 75% * $200 = $150, which is more than $100, implying that a risky choice should be favored). In more naturalistic cases, whether evaluating a credit card offer or the personal advice of a friend, the logic is functionally the same. Rational, normatively superior decisions may be defined by optimization analyses that coherently integrate values, goals, preferences, and constraints in accord with *standards of logic, probability, and statistics* (Baron, 1985, 2008; Edwards, 1954; but see Gigerenzer, Todd, & the ABC Research Group, 1999).

It is noteworthy how well most people's decisions approximate various normative standards, *as if* they actually solved an econometric or Bayesian statistical equation (Chater, Tenenbaum, & Yuille, 2006). However, beyond shared conceptual elements, optimization methods bear little resemblance to actual human decision processes: most people do not compute statistical analyses in their head for the hundreds of decisions they make every day (e.g. selecting shoes or breakfast cereals, deciding who to talk with and what to discuss). In many naturalistic contexts decisions actually entail so much complexity and uncertainty that comprehensive optimization is impossible even for the most powerful computers. Because people have limited time, knowledge, and cognitive resources, alternative decision strategies are required, leading people to rely on the simple heuristics that so often empower effective decision-making (e.g. satisficing, take-the-best, recognition, fluency; Gigerenzer et al., 1999; Klein, 1999; Simon, 1956, 1990). Taken together, the literature reviewed here indicates (a) that statistical numeracy robustly predicts general decision-making skill and (b) that skilled decision-making is fundamentally about reckoning with risk and uncertainty. However, (c) the link between numeracy and superior decision-making cannot be explained by differences in the use of formal optimization methods because (d) most decisions are ill-structured, making them so complex that optimization is practically impossible. Thus, a central question remains: What are the causal mechanisms that give rise to normatively superior human decision-making? To address this question, we turn to research on skilled decision-making in experts and non-experts.

Expert and Skilled Decision-Making

Human cognition, including analytical and emotional processing, is often characterized with respect to the interplay of intuition and deliberation (e.g. automatic versus controlled, fast versus slow thinking, hot and emotional versus cold and calculating, dual-system, dual-types, and many others; Evans & Frankish, 2009; Kahneman, 2003, 2011; Stanovich & West, 2000; Shiffrin & Schneider, 1977; but also see Arkes, 2016; Cokely, 2009; Moshman, 2000; Osman, 2004). Common assumptions suggest that superior decision-making may generally require outstanding cognitive abilities that enable the inhibition of emotions and intuitions, while empowering complex and abstract logical reasoning. This perspective offers something like a *Mr. Spock* depiction of the pinnacle of skilled human decision-making (i.e. the extremely logical and emotionless science-fiction superintellect). Clearly, some professionals do routinely perform technical or formal decision analyses for high-stakes decision-making (e.g. operations research). Computerized decision support technologies are also increasingly used to inform diverse decisions (e.g. forecasting, medical decision-making). However, beyond these few examples, research reveals a very different picture of the underlying cognitive dynamics involved in skilled decision-making, as extensively documented in the scientific literature on human expert performance (for state-of-the-science reviews see Ericsson, 1991; Ericsson, Charness, Hoffman, & Feltovich, 2006; Ericsson, Prietula, & Cokely, 2007; and this volume).

Expert Performance

The term *expert* colloquially refers to trained professionals with experience or credentials (e.g. a qualified expert). In contrast, *expert performance* is a scientific term referring to the verifiable, reproducible, and superior human performance that can be exhibited (and studied) in representative and naturalistic tasks. The magnitude of the superiority exhibited by expert performers tends to be remarkable (e.g. 3 to 1000+ standard deviations better as compared to novices). For example, many chess grand masters can simultaneously play hundreds of games, readily beating nearly all their highly skilled and motivated competitors (e.g. the world record involved 604 simultaneous matches with only 8 losses). Research on these kinds of remarkable abilities, based on the expert performance approach, has revealed a great deal about the mechanisms that mediate and govern verifiable expertise. For example, no one ever becomes an expert in an established domain of expert performance without first deliberately practicing for thousands of hours over many years (Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson, Prietula, & Cokely, 2007). Deliberate practice specifically refers to specialized, high concentration training efforts, often completed in solitude. These efforts are intentionally designed to move the expert past their current performance level through elicitation of specific feedback about performance weaknesses and strengths. Accordingly, deliberate practice tends to be quite challenging and cannot be routinely sus- tained for great periods of time (e.g. perhaps four to five hours per day on average). Unlike playing golf for fun with friends, one might delib- erately practice by independently working on a very specific kind of bunker shot under controlled conditions. In time, careful and consistent deliberate practice efforts tend to produce remarkable differences in knowledge, skills, and abilities.

Expert Decision-Making

Studies of chess expertise efficiently illustrate the cognitive mechanisms that support human expert decision-making more generally (for a review see Gobet & Charness, Chapter 31, this volume). Decision research in chess is facilitated because the rules are well specified and the goals are unambiguous. The game has a long tradition with formal scoring conventions that index spe- cific ratings for a huge number of past and present players. Chess is also computationally complex: there is currently no known solution or optimal strategy that can ensure a victory regardless of an opponent's move. Thus, consistently superior decision-making in chess involves heuristic deliberation and logical inductive reasoning under risk and uncertainty, in accord with formal foundations of normative decision and rational choice theories. As such, chess computer programs attempt to approximate an optimal solution by searching hundreds of millions of positions per second using heuristics to estimate the relative expected values of candidate moves. In contrast, extended search and formal computation play an exceedingly small role among human experts. Instead of selecting moves based on massive iterative search processes, verifiable experts rely on their vast and integrated stores of specialized knowledge in long-term memory, facilitating rapid encoding of goal-relevant features in tandem with sophisticated and nuanced pattern recognition. This integrated understanding allows experts to quickly narrow their search and evaluation so they can deliberatively evaluate a small number of outstanding candidate options (e.g. typically around four per minute).

The comparison of chess experts to less skilled chess players reveals several notable skill differences. Experts can extract useful patterns and relations from information faster and in parallel, while less skilled individuals rely on slower serial encoding and search. In naturalistic non-routine situations experts tend to deliberate more during move selection compared to less skilled individuals, and this difference tends to benefit their performance (Moxley, Ericsson, Charness, & Krampe, 2012). Most importantly, experts consider qualitatively different moves than those considered by less skilled individuals. This reflects substantial differences in their fundamental understanding and knowledge of game situations: expert decision-makers have access to more and more sophisticated chunks in long-term memory (e.g. have memorized more than 100,000 moves and sequences). Ultimately, experts rely on larger intricately structured knowledge databases in memory to support and evaluate a small number of complex inductive mental simulations. Taken together, these findings highlight the power and primacy of knowledge and sophisticated understanding over formal computation in verifiable expert decision-making (e.g. Ericsson et al., 2006; Klein, 1999).

General Decision-Making Skill

Given that even experts don't make decisions by trying to imitate formal mathematical optimization methods, perhaps it is not surprising that skilled, non-expert decision-makers also typically forgo formal calculation across wide-ranging everyday decision contexts. This pattern is well illustrated in the first cognitive process tracing study to directly map the links between decision strategies, cognitive abilities, and superior decision-making under risk (Cokely & Kelley, 2009). Using choice outcome modeling, together with decision reaction time and retrospective verbal protocol analysis (Ericsson & Simon, 1980, 1993; Fox, Ericsson, & Best, 2011), the study mapped the strategies that individuals with higher cognitive ability scores (i.e. working memory, numeracy, and cognitive reflection) used to make superior decisions when tasked with paradigmatic risky prospect evaluation. Analyses revealed that only a very small proportion of people attempted to explicitly calculate expected values (i.e. < 5 percent explicitly multiplied probabilities by values). Instead, the vast majority of superior decision-making was found to be linked to personally relevant, affectively charged heuristic-based evaluation. Structural modeling further revealed that all general ability-to-performance relations were fully mediated by large differences in affective and elaborative long-term memory encoding and metacognitive reasoning (i.e. evaluating relations between feelings, thoughts, and consequences in personally relevant concrete mental simulations and narratives).

Results from Cokely and Kelley (2009) suggested that even during ultra-simplified paradigmatic risky prospect evaluation, general decision-making skill tends to reflect differences in how and how much people think about, and meaningfully understand, a decision problem (see also Ghazal et al., 2014; Jasper, Bhattacharya, & Corser, 2017; Pachur & Galesic, 2013; Peters, 2012, 2017; Reyna et al., 2009). Rather than calculating expected outcomes, most skilled decision-makers spent more time imagining how changes in wealth would affect their lives and how those changes might feel via informal narratives (e.g. "well that's probably never going to happen but if it did how could I forgive myself"). Indeed, some of the least cognitively "able" individuals were among the most skilled decision-makers, reflecting their more extensive, personally meaningful deliberation. This sort of active, careful, and open-minded exploration accords with theories of successful decision-making as detailed in Baron's touchstone work *Rationality and Intelligence* (1985). Today many related studies have been conducted in our labs and others, including one study with more than 50,000 participants from 46 different countries (Rubinstein, 2013). Additional converging process tracing evidence comes from protocol analyses, eye-tracking, reaction time analyses, choice modeling, memory analyses, and causal experimentation (Garcia-Retamero & Cokely, 2013, 2014, 2017; Garcia-Retamero, Cokely, Wicki, & Joeris, 2016b; Ghazal et al., 2014; Okan, Garcia-Retamero, Cokely, & Maldonado, 2015; Woller-Carter, Okan, Cokely, & Garcia-Retamero, 2012). Taken together, results suggest that general decision-making skill often involves and can be predicted by both quantitative and qualitative differences in heuristic-based deliberation and representative understanding in long-term memory. These findings contribute to an integrative theoretical account of skilled human decision-making.

Skilled Decision Theory

Skilled Decision Theory explains the essential causal mechanisms that enable skilled decision-making in experts and non-experts with reference to the pivotal roles of heuristic deliberation and representative understanding. Skilled Decision Theory is grounded in Skilled Memory Theory (Ericsson, Chase, & Faloon, 1980), extending our understanding of expert decision-making in accord with frameworks for *rational thinking* (Baron, 1985, 2008) and *adaptive heuristic decision-making* (Gigerenzer et al., 1999). The causal mechanisms that support general decision-making skill are similar to those that support complex *situation model* development in skilled reading comprehension (i.e. an acquired domain-general skill) and high-fidelity *situation awareness* in expert decision-making (i.e. an acquired domain-specific expertise). Because most adults have a vast expert-like knowledge of themselves (e.g. experiences, values, habits, goals, desires, and preferences) personally meaningful heuristic deliberation facilitates rapid detection and detailed encoding of relevant information in durable, integrated long-term memory representations (e.g. mental models, situation narratives, and inductive simulations; Holland et al., 1986). By utilizing extensive and thoroughly integrated pre-existing knowledge structures, non-expert decision-makers functionally circumvent attentional capacity limitations of short-term (working) memory that could otherwise constrain the complexity and precision of their understanding and reasoning, thereby leveraging the same kinds of long-term working memory capacities that support expert performance more generally (Ericsson & Kinstch, 1995).

Despite substantial benefits, personally oriented heuristic deliberation can increase the likelihood of potentially counterproductive processes (e.g. confirmation bias, anchoring, attribute substitution, biased sampling, small sample overgeneralization, affective overshadowing, overweighting priors, and others). In part, this is why statistical numeracy and associated metacognitive skills are essential components of general decision-making skill. Decision-makers who are skilled in *practical inductive reasoning* (e.g. statistically numerate and metacognitively savvy), and who also engage in elaborative heuristic deliberation, are well prepared to correct or circumvent potentially costly mistakes via metacognitive heuristics (e.g. disconfirming, reframing, resampling, double checking, base rate conditioning, affect recalibrating, and coherence checking; Cokely et al., 2012; Ghazal et al., 2014; see also Baron, 1985, 2008). Thus, ordinary people are able to generate a detailed, relatively coherent, and representative understanding of the decision problem allowing them to intuitively yet precisely *feel the weight and potential consequences* of various options and outcomes (Peters, 2012; Petrova, Garcia-Retamero, & Cokely, 2015; Petrova, van der Pligt, & Garcia-Retamero, 2014; Petrova et al., 2017a; Traczyk & Fulawka, 2016; but also see *Fuzzy-Trace Theory* for a perspective emphasizing benefits of imprecise, automatic aspects of comprehension; Reyna, 2004, 2008; Reyna & Brainerd, 1995). In turn, this understanding informs the selection of adaptive heuristic strategies such that decisions can approximate normative standards without any formal optimization analyses (e.g. ecologically rational representation of cue validities, cue orderings, and other factors that calibrate fast and frugal heuristic use; Gigerenzer, 2015; Gigerenzer & Goldstein, 1996). In these and other ways, essential skills promote resilient and adaptive decision-making under conditions of complex risk and uncertainty, without requiring or guaranteeing neo-classical rationality.

In summary, among typical adults, including non-experts and experts alike, Skilled Decision Theory holds that superior decision-making is primarily driven by (i) skilled and personally relevant heuristic deliberation and (ii) sophisticated, affectively charged representative understanding that (iii) interacts with and informs adaptive heuristic use (Cokely et al., 2012; Cokely, Schooler, & Gigerenzer, 2009; Garcia-Retamero & Cokely, 2017; Gigerenzer & Gaissmaier, 2011; Keller, Cokely, Katsikopoulos, & Wegwarth, 2010). Taken altogether, the observed primary roles of acquired knowledge and heuristic deliberation sharply contrast with longstanding assumptions about the importance of abstract deductive logical reasoning capacities. To help resolve this apparent inconsistency, we next review recent and major psychometric studies of general cognitive abilities.

General Intelligence and Decision-Making Skill

In 1994 a controversial book was published entitled *The Bell Curve: Intelligence and Class Structure in American Life* (Herrnstein & Murray, 1994). This book was presented as a crowning synthesis of more than a century of research and theory on individual differences in intelligence. The book also included a new major study of links between general intelligence and life outcomes, suggesting wide-ranging potential implications, broadly arguing that: (a) general intelligence was one of the most influential *causes* of economic and social prosperity in the United States (and probably elsewhere) and (b) intelligence was largely but not entirely determined by genetic factors (e.g. roughly stable with firm upper-bounding individual limits). Ultimately, the book detailed a common perspective suggesting that the primary reason intelligence predicts life outcomes is because intelligence tends to be a strong determinant of decision-making quality. According to this view, only a small number of people have the rare intellectual aptitudes that are required for consistently good decision-making ("cognitive elite"), given the increasing social and technological complexities of our industrialized world. Indeed, a huge and orderly body of data documents a robust, albeit often modest, link between general intelligence and life outcomes including educational attainment, wealth, occupational achievement, professional advancement, health, and others. However, other claims remain largely unsubstantiated, including (a) the causal assumptions about the underlying mechanisms that allow general intelligence to predict life outcomes (e.g. fixed intellectual capacities), (b) the extrapolations about perceived intellectual trends (e.g. the emergence and dominance of a ruling class of cognitive elite), and (c) the appropriate policy prescriptions (e.g. social, educational, and economic segregation).

Although links between intelligence and outcomes have been well established for nearly a century, until recently direct evidence on the connections between basic cognitive abilities, decision-making, and life outcomes has been extremely limited. Consider what is regarded as the most comprehensive integrative analysis of the structure of general cognitive abilities and intelligence to date, namely the book entitled *Human Cognitive Abilities : A Survey of Factor-Analytic Studies* by John Carroll (1993). This monumental monograph presents a reanalysis of 460 factor-analytic cognitive ability studies collected over a 60-year period. Across all included datasets and tasks *decision-making* was rarely mentioned, except in the context of very simple perceptual reaction-speed tasks. However, tests of reasoning abilities received extensive attention and featured prominently in Carroll's estimates of the fluid intelligence factor, which was determined to be the strongest single factor explaining overall general intelligence (Figure 26.1).

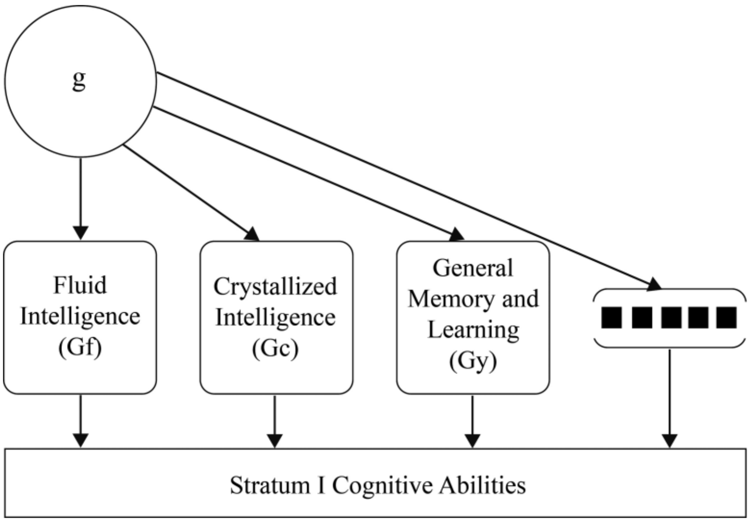

Figure 26.1 The total variance explained in overall general decision-making skill by each variable modeled as a single, sole predictor.

Based on analyses of 176 large datasets including some 236 different reasoning factors Carroll (1993) suggested that three essential reasoning factors best explained fluid intelligence, namely (1) sequential reasoning, (2) inductive factors, and (3) quantitative reasoning. Careful inspection reveals that paradigmatic decision-making skills were nevertheless broadly neglected. In contrast to common assumptions, the category labeled "inductive factors" does not broadly represent inductive *logic*. The half-dozen or so inductive tasks instead primarily require narrow types of inferences used during abstract deductive reasoning. Specifically, within deductive mathematical reasoning there is a class of "inductive inference" that is relatively independent of risk or uncertainty. In psychological research these kinds of deductive reasoning tasks – including progressive matrices, analogical reasoning, and categorization-type problem-solving – are more specifically referred to as "rule induction" tasks (Goldman & Pellegrino, 1984; Holyoak & Morrison, 2005). Across all these types of tasks, if a person can figure out the rule that precisely dictates the relations among various elements, they can *always* deduce the final answer with absolute certainty.

To be clear, there is surely some psychological uncertainty involved in the completion of deductive rule-induction tasks (e.g. matrix reasoning), such as deciding when you should be confident you have determined the correct rule before deducing a final answer. However, these tasks are otherwise not indices of the full range of inductive logic that is characteristic of modern paradigmatic or naturalistic decision-making. Moreover, the "quantitative reasoning" tasks used to measure fluid intelligence also typically failed to assess the full range of relevant skills, focusing primarily on conventional mathematical components that "require an appreciation of the quantitative concepts and relations, particularly as treated in mathematics in its various branches, from simple arithmetic to algebra, geometry, and calculus" (Carroll, 1993, p. 213). As a result of this narrow task-sampling, most major factor-analytic intelligence studies conducted in the twentieth century fail to incorporate a wide range of logical inductive, probabilistic, or statis- tical reasoning skills in their analyses.

Historical coincidence partially explains why decision science perspectives have been neglected in psychometric intelligence research. Behavioral decision science is a relatively young field that began to have a mainstream presence in psychology and economics about 25 years after the cognitive revolution of the 1950s. Statistical theory and training is also a relatively modern invention as the axiomatic (objective) formalization of probability theory was not published until 1933, about two years before Ronald Fisher inspired a scientific revolution with *The Design of Experiments*. And still today highly trained and well-respected scientists often misunderstand and misapply statistical theory (Lindsay, 2015). Moreover, even conventional quantitative skills have historically been treated inconsistently in the intelligence literature. For example, Herrnstein and Murray (1994) excluded a brief (five minute) timed test of mathematical operations from their final analysis because it was not sufficiently correlated with their intelligence estimate (i.e. the data were available but excluded due to low factor loadings). Reanalyses of their data found that the excluded timed math operations test actually predicted wages as well as the part of the Armed Forces Qualifications Test data that Herrnstein and Murray used (Heckman, 1995). If indeed this very brief test of math ability was relatively independent of "general intelligence," yet still predicted wages as well as the unrelated estimates of general intelligence, it follows that numerical abilities must explain important life outcomes that are otherwise missed by more standard general intelligence metrics. To evaluate the relative influence of intelligence and other cognitive abilities on general decision-making skill, we next consider several recent landmark psychometric decision science studies.

Psychometric Studies of Decision-Making Skill

At the turn of this century, Stanovich and West (2000) published a highly influential paper entitled "Advancing the Rationality Debate." Their work reviewed individual differences in decision-making, examining the link between general cognitive abilities and susceptibility to common heuristics and biases in the Kahneman and Tversky tradition (Stanovich, 1999). Based on analyses of correlations, Stanovich and West showed that intelligence predicted normatively superior decision-making in young adult participants, suggesting (1) a positive manifold among many judgment and decision-making abilities, and (2) a robust correlation with general intelligence as measured by standardized college achievement or matrix reasoning tests (i.e. deductive rule-induction tasks). Although their theoretical account of these findings has changed in fundamental ways over the last 15 years (Stanovich, 2016; Stanovich, West, & Toplak, 2016), the groundbreaking work of Stanovich, West, and others paved the way for more comprehensive psychometric decision science investigations (see also the seminal contributions of Frederick, 2005; Peters et al., 2006).

Decision Competency Assessment. Historically, research has emphasized four core decision skill components known to shape decision *consistency* and *accuracy*, namely (1) assessing beliefs, (2) assessing values, (3) integrating beliefs and values, and (4) having a metacognitive understanding of one's abilities, resources, and constraints (Edwards, 1954; Raiffa, 1968; Yates, 1990). Leveraging this reasoning, in 2005 the first in a series of two landmark studies presented what was then the most comprehensive assessment of "plausible real-world correlates of good decision making … [that broadly] span the domain of decision-making skills" (Parker & Fischhoff, 2005, p. 1). Specifically, the study assessed and modeled general decision performance and ability structure across seven paradigmatic decision domains, namely, (1) risk perception consistency, (2) social norm recognition, (3) resistance to sunk costs, (4) resistance to framing, (5) applying decision rules, (6), choice path independence/consistency, and (7) confidence calibration (e.g. over-/under-confidence). Integrating the patchwork of individual difference assessments used in the past, Parker and Fischoff created a broad Young Decision Making Competency assessment (Y-DMC) and administered it to 110 diverse young adults.

Parker and Fischhoff's (2005) study revealed an anticipated positive manifold among paradigmatic decision tasks, which was well explained by a one-factor model. Analyses showed that the Y-DMC was related to other cognitive ability tests (e.g. vocabulary and executive control). Importantly, each of the seven facets also predicted various behavioral outcomes, decision styles, or risk factors ranging from sustaining social support and positive peer interaction, to delinquency, drug use, and other vulnerabilities. These relations remained significant and largely unchanged after statistically controlling for general cognitive abilities. Results further indicated that decision-making skill largely or entirely mediated the link between general cognitive abilities and important life outcomes such as behavioral coping, externalizing behaviors, and overall at-risk status.

Two years after the initial validation, a second landmark study in this series was published by Bruine de Bruin, Parker, and Fischhoff (2007). The research included an updated, advanced competence assessment designed for use with diverse adults from industrialized countries (i.e. the A-DMC; Table 26.1). The extensive decision-making assessment was completed at home by 360 participants. Once again, a one-factor model provided an efficient fit, although a two-factor model improved model fit (e.g. from 30 percent to 46 percent explained variance). Predictive validity was demonstrated using a Decision

Table 26.1 Components of the Adult Decision-Making Competence (A-DMC) assessment (Bruine de Bruin et al., 2007).

A-DMC component	Description	Example item
Consistency in Risk Perception	This task asks participants to judge the probability of various events happening in two different time frames.	probability that someone will steal something from you during the next year/in the next 5 years?
Recognizing Social Norms	This test measures how well participants judge social norms. Participants assess 16 undesirable behaviors.	First set: "It is sometimes OK to steal under certain circumstances." Second set: "Out of 100 people your age, how many would say it is sometimes OK to steal under certain circumstances."
Resistance to Sunk Costs	This test measures the ability to ignore prior financial and time investments that are no longer paying off when making decisions.	After a large meal at a restaurant, you order a big dessert with chocolate and ice cream. After a few bites you find you are full and you would rather not eat any more of it. Would you be more likely to eat more or to stop eating it?
Resistance to Framing	This task measures whether value judgments are affected by irrelevant variations in how the problem is presented.	Recent evidence has shown that a pesticide is threatening the lives of 1,200 animals. Two response options have been suggested. Which option do you recommend: (1) Option A: 600 animals will be lost for sure. (2) Option B: 75% chance 400 animals will be lost, and 25% chance that 1,200 animals will be lost. The same item is then presented in a "gain" format (e.g. 600 animals are saved for sure).
Applying Decision Rules	This task evaluates the ability to apply decision rules, by asking participants to choose between DVD players with different ratings and features.	Lisa wants the DVD player with the highest average rating across features. Which one of the presented DVD players would Lisa prefer?
Path Independence	This test presents item pairs posing normatively equivalent choices between gambles. The participants' choice should not be affected by normatively irrelevant changes.	Which do you like best: (1) Flip a coin. If heads, win $100. If tails, win $0. (2) Sure Win. Win $50 for sure. Also presented in different forms. Performance measured by participant's consistency in choices.
Under-/over-confidence	This test measures how well participants can assess their own knowledge. Participants first answer a true/false question, then assess their confidence in that answer.What is the	True or false: Stress makes it easier to form bad habits. How confident are you? 50% (just guessing) to 100% (absolutely sure).

Outcomes Inventory (DOI), a robust self-report general weighted index of maladaptive real-world decision outcomes (e.g. being fined, overspending). General intelligence was also directly assessed using the Raven's advanced progressive matrix reasoning test (fluid intelligence) and the Nelson-Denny reading test of verbal ability (crystallized intelligence).

The results of Bruine de Bruin et al. (2007) revealed strong relations between intelligence and general decision-making skill (i.e. correlations around 0.6). However, predictive validity modeling showed the composite A-DMC decision quality score predicted decision outcomes about four times better than intelligence scores alone. Decision-making skill as assessed by A-DMC also explained about 75 percent of the link between intelligence and decision outcomes, as measured by the DOI. This finding suggests that a primary reason intelligence predicts better decision outcomes is because more intelligent people tend to acquire higher levels of decision-making skill. Nevertheless, estimated decision-making skill scores predicted decision outcomes much better than, and largely independently of, intelligence (i.e. 80 percent of decision-making skill's predictive power was independent of intelligence). Taken together, these landmark studies suggest that while standard intelligence tests predict real-world decision-making outcomes, they tend to be much less powerful compared to more direct measures of general decision-making skill. Results further confirm that paradigmatic decision-making competency provides a robust estimate of real-world decision-making skill, predicting high-stakes decision outcomes far better than, and largely independent of, general intelligence (Bruine de Bruin et al., 2007; Del Missier, Mäntylä, & Bruin, 2012; Parker & Fischhoff, 2005).

Intelligence, Decisions, and Numeracy Components The work of Parker and Fischhoff (2005) and Bruine de Bruin et al. (2007) provided a robust and representative benchmark assessment of paradigmatic decision-making skills including (but not limited to) tasks in the heuristics and biases tradition (Kahneman, 2011; Tversky & Kahneman, 1985). However, until recently, broader efficient research assessments covering the full range of quantitative numeracy skills have not been widely available. To address this gap we turn to a comprehensive adult numeracy framework (Ginsburg et al., 2006) derived following a systematic review of 29 existing mathematical and numeracy frameworks and related national education standards. This framework indicates that the modern *core* collection of essential components of adult numeracy in industrialized societies typically involves: *operations* including computation, estimation, rates, ratios, proportions, and percentages; *probability* including collection, organization, and display of data, analysis and interpretation of data, chance and probability, and inferential reasoning; *geometry* including measurement units, trigonometric ratios, angles and lines, shapes, perimeter, area, and volume, length, width, height, and radius; and *algebra* including algebraic expressions, symbols, equations, and functions. Building on this and related statistical literacy frameworks (Gal, 2003; Ginsburg et al., 2006; Kutner et al., 2006), we developed the Berlin Numeracy Components Tests (BNT-C), using a multiphase iterative test development process, in order to provide simultaneous estimates of:

1. Full-scale adult numeracy.
2. Adult numeracy subscales (i.e. operations, probability, algebra, geometry).
3. Statistical numeracy (i.e. a composite of operations and probabilities).
4. Conventional numeracy (i.e. a composite of algebra and geometry).

Several two-parameter logistic models were estimated using Item Response Theory in order to identify a final pool of (up to) 36 items for efficient assessment across various samples and subsamples of diverse people in industrialized countries (e.g. all test/sub-test information functions peaked around theta of zero; Cokely, Allan, Ghazal, Feltz, & Garcia-Retamero, forthcoming; Ghazal, 2014). As with previous tests, different formats allowed items to be presented in adaptive or traditional modes (e.g. paper and pencil), as needed for various study constraints (e.g. brief online versus extensive in-person testing). Thanks in part to essential funding from the National Science Foundation and others, we then endeavored to conduct the most comprehensive and representative series of studies of general decision-making skill in history, while simultaneously updating national and international norms for numeracy, risk literacy, and decision-making skill (Cokely et al., forthcoming; but for similar extensive and ambitious ongoing approaches that have also revealed strong and robust influences of numeracy see Stanovich et al., 2016).

In one of our first broad studies, we included hundreds of paradigmatic and ecologically sampled decision-making tasks (see Table 26.2), including (a) the Adult Decision-Making Competence assessment (A-DMC), (b) a paradigmatic prospect evaluation assessment battery (e.g. risky lotteries, intertemporal choice), (c) a class inclusion illusion task battery (e.g. denominator neglect, ratio bias), and (d) an ecological risk literacy test battery made up of real-world decision and evaluation tasks sampled from representative health, financial, natural hazard, civic, and consumer contexts (e.g. evaluating real advertising and political polls, interpreting relevant medical risks, offering actual recommendation to peers about HIV, relationships, finances, etc.). One hundred and twenty-six young adults completed all phases of the assessment including all decision-making tasks, the newly developed Berlin Numeracy Components Test (BNT-C), seven other leading numeracy tests, the Raven's advanced progressive matrices for assessment of fluid intelligence, the assessment of cognitive impulsivity developed by Frederick (2005), as well as about two dozen other personality, trait, style, ability, demographic, and outcome assessments.

Upon completion of the study we derived an overall general decision-making skill estimate based on a weighted cumulative index of superior decision-making performance (e.g. weighted by relative factor loadings). As measured with the Berlin Numeracy Components Test alone, full scale numeracy accounted for 34 percent of the total variance in overall decision-making skill. In comparison, the best combination of all other cognitive ability and numeracy instruments accounted for roughly 30 percent of overall decision-making skill, as depicted in Figure 26.2. Despite taking more than ten times longer to complete, all other cognitive ability tests combined provided significantly less predictive power compared to the single brief Berlin Numeracy Components Test. Analyses further indicated that the statistical numeracy sub-test portion alone explained 33 percent of the total decision-making skill variance (as a single predictor), such that 97 percent of the predictive power of full-scale numeracy was shared with the statistical numeracy subscale. This finding highlights the widely observed, robust link between general decision-making skill and statistical numeracy (e.g. a five-minute test specifically focusing on operations and probabilistic inductive reasoning skills). By nearly any behavioral science standard this association is very strong. It is similar to the relationship between temperature and distance from the equator in the United States (e.g. Michigan is almost always much colder than Oklahoma or Florida), or more than 25 times greater than the meta-analytically derived average effect of ibuprofen for acute pain relief. In comparison, the predictive power of conventional numeracy alone (i.e. geometry and algebra) explained less than half as much variance as statistical numeracy (i.e. 16 percent). To illustrate, Figure 26.2 presents data from a sample of young adults displaying the total

Table 26.2 General decision-making skill assessment battery.

Component	Description	Example item
A-DMC	Includes a battery of seven categories of paradigmatic decision tasks and common biases.	See Table 26.1
Prospect Evaluation		
Risky prospects and lotteries	Includes paradigmatic, modest-stakes risky prospect evaluations presented in many formats, across a wide range of risk and EV ratio ranges.	Which of the two options do you prefer: (1) Lose $50. (2) 50% chance to lose $400.
Intertemporal choices	Presented a series of paradigmatic time-reward preference tasks based on previous research. Items present differing time intervals and reward amounts.	Which of the two options do you prefer: (1) $3400 this month. (2) $3800 next month.
Reference Class and Class-Inclusion Neglect, with confidence calibration	Measures the propensity to neglect reference classes, base rates, and various ratio-relevant decision factors (e.g. denominators).	A new drug is introduced to reduce the risk of death from a heart attack for people with high cholesterol. A study of 900 people with high cholesterol showed that 80 of the 800 people who have not taken the drug died after a heart attack, compared with 16 of the 100 people who did take the drug. How beneficial was the drug? (plus confidence)
Ecological Risk Literacy and Informed Decision-Making, with confidence calibration	Assesses informed decision-making in risky health, financial, and natural hazard decisions sampled from, and ecologically representative of, common naturalistic decision tasks.	Imagine you take out a $50,000 federal student loan to help pay for college. You are offered four possible repayment plans. Given information about each of these plans, participants must determine which plan is best, based on dif- ferent criteria. (plus confidence)

variance explained in overall superior decision-making performance (i.e. general decision-making skill) by each of the following single predictors alone: *Full-Scale Numeracy* (i.e. statistical numeracy + conventional numeracy), *Statistical Numeracy* (i.e. probability + operations), all other assessed *General Cognitive Abilities* (e.g. fluid intelligence, cognitive reflection, health literacy, other numeracy tests), *Conventional Numeracy* (i.e. algebra + geome- try), and *Raven's Advanced Progressive Matrices* (i.e. fluid intelligence) (Cokely et al., forthcom- ing; Ghazal, 2014).

Additional analyses confirmed that numeracy was strongly related to fluid intelligence in our study (i.e. about 25 percent shared variance), with

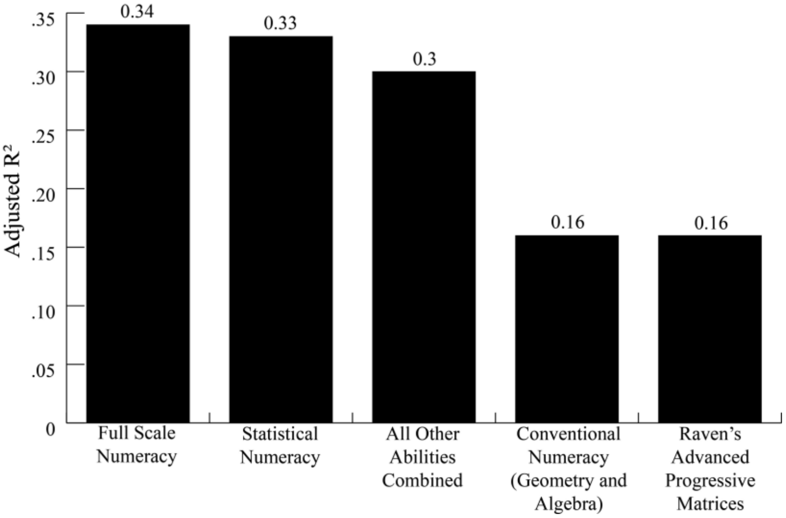

Figure 26.2 The total variance explained in overall general decision-making skill by each variable modeled as a single, sole predictor.

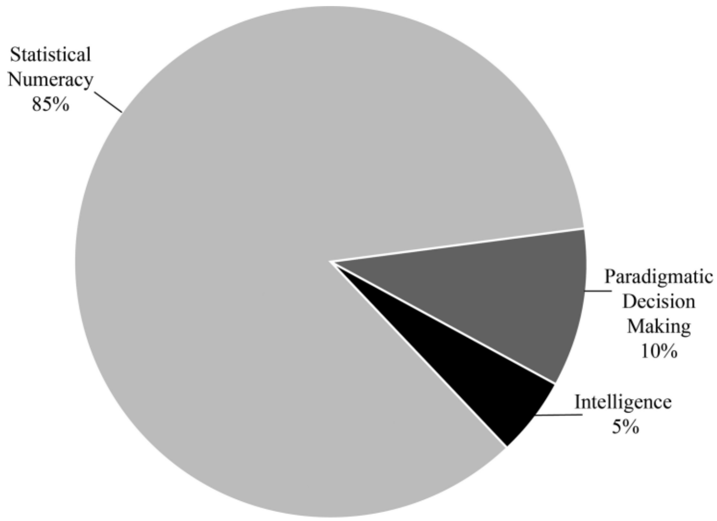

Figure 26.3 The proportion of variance in a factor-analytically derived best-fitting model of overall general decision-making skill.

fluid intelligence exhibiting a significantly stronger association with conventional numeracy than statistical numeracy. Fluid intelligence alone (as assessed by the Raven's advanced pro- gressive matrices) explained less than half the variance of the brief statistical numeracy test, as is depicted in Figures 26.3–26.5. Structural models further indicated that the link between fluid intelligence and general decision-making skill was largely mediated by statistical numeracy. Consistent with previous studies, fluid intelligence predicted general decision- making skill in large part because more intelligent people had higher levels of statistical numeracy proficiency. That said, high levels of statistical numeracy and decision-making skill also occurred in the absence of high fluid intelligence scores. Similar patterns of results were found in fine-grained analyses of specific decision competencies (e.g. A-DMC, Ecological Risk Literacy, Class-Inclusion Illusions, and Risky Prospect Evaluation). In all cases the strongest single predictor of decision-making skill was full-scale numeracy, followed closely by statistical numeracy, which almost entirely mediated any connection between fluid intelligence and decision-making skill. In turn, subsequent structural modeling revealed that one's ability to evaluate and understand risk (i.e. risk literacy) tended to mediate the observed relations between statistical numeracy and general decision-making performance: About 70 percent of the relationship between numeracy and decision-making was explained by decision-making performance on naturalistic risky decision-making tasks (e.g. assessing real financial products, evaluating treatment advice, interpreting risks and trade-offs based on actual scientific evidence or government-approved brochures, etc.). These results converge, indicating that acquired statistical

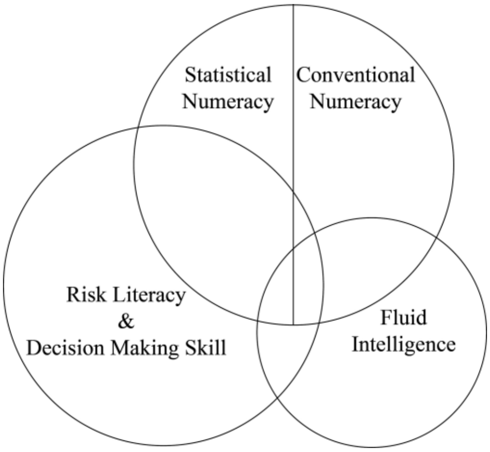

Figure 26.4 Spatial depiction of the estimated and approximate scaled relations between (a) statistical numeracy, (b) conventional numeracy, (c) risk literacy and general decision-making skill, and (d) fluid intelligence.

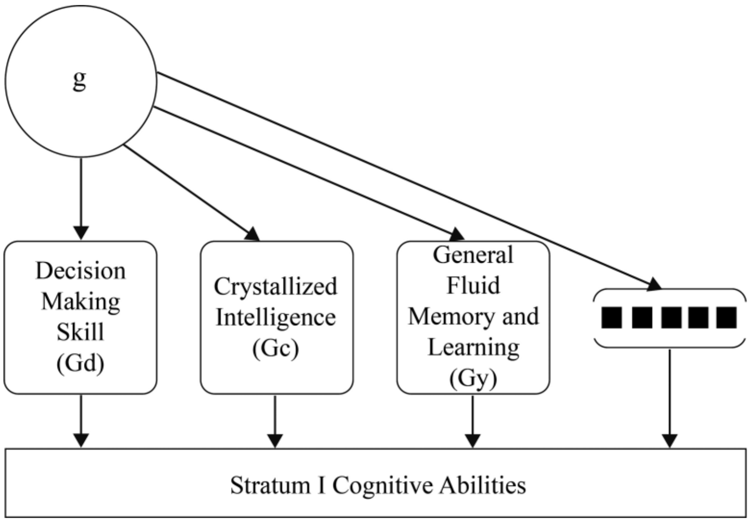

Figure 26.5 The predicted hypothetical restructuring of Carroll's (1993) cognitive ability model based on emerging data, highlighting the robust influence of general decision-making skill.

numeracy skills robustly and uniquely predict one's general decision-making skill, often operating independent of fluid intelligence and other basic cognitive abilities among diverse and generally healthy young adults. To further illustrate these relations, based on a sample of young adults, Figure 26.3 depicts the proportion of variance in a factor-analytically derived best-fitting model of overall *general decision-making skill* as a function of (a) Statistical Numeracy, (b) Paradigmatic Decision-Making Tasks (e.g. risky prospect evaluation), and (c) Fluid Intelligence (i.e. Raven's Advanced Progressive Matrices).

How Numeracy Out-Predicts Fluid Intelligence

The tight connections linking risk, uncertainty, and decision-making help explain why statistical numeracy is such a strong predictor of decision-making skill: decision-making is fundamentally about reckoning with risk and uncertainty (Figure 26.4). Accordingly, a normatively super- ior decision is essentially a good bet. If you know the rules of probability theory and understand your own information processing biases, competencies, knowledge, and values, no matter how carefully you check your reasoning, your choice will *always entail some risk and uncertainty*. In contrast, fluid intelligence tests are primarily about careful and thorough deductive reasoning under conditions of certainty. Thus, if you carefully concentrate for long enough to triple check the coherence of your analysis, you can determine the right answers on intelligence tests with *perfect certainty every time*.

The fact that abstract deductive reasoning (i.e. rule-induction) is a central feature of "gold standard" fluid intelligence tests helps explain why these tests are argued to be best characterized as assessments of working memory and attentional control (e.g. general fluid memory; Kyllonen & Christal, 1990; McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). One's ability to coherently sustain and direct high-concentration attention in the service of error-free analysis is especially relevant for time-limited problem-solving tasks involving complicated abstract rules and unfamiliar artificial objects. For example, a worry-free person who wakes up fresh in the morning might answer nearly every matrix question correctly, yet that same person may cut corners and make strange processing and logical errors if sleep deprived, hungry, upset, or otherwise preoccupied (e.g. lapses of attention). Indeed, many kinds of acquired, environmental, motivational, and emotional mechanisms influence these attentional control capacities, as reflected in the very large historical increases in fluid intelligence scores observed during the twentieth century (Fox & Mitchum, 2013, 2014; but for training and motivation studies see Cokely, Kelley, & Gilchrist, 2006; Duckworth & Seligman, 2005; Jaeggi, Buschkuehl, Jonides, & Shah, 2011).

Because fluid intelligence tests are fundamentally tests of abstract (unfamiliar) reasoning under certainty, the task demands imposed by fluid intelligence tests are not broadly representative of typical human decision-making (i.e. not ecologically representative or valid). In contrast, actual human decision-making is largely a knowledge-centric and comprehension-oriented activity. Effective decision-making generally involves consideration of options in the context of extensive knowledge, deeply held values, actual responsibilities, and practical constraints (e.g. with reference to and relevance for family, finances, health, career, happiness, trust, regret, safety, etc.). The precise evaluation of costs, benefits, and trade-offs necessarily entails concrete consideration of familiar, tangible, and affectively charged outcomes and implications (e.g. losing money, avoiding illness, giving risky advice). In some sense, trying to measure decision-making skill with an unfamiliar abstract rule-induction task may be like trying to measure reading comprehension with an unfamiliar foreign vocabulary learning test. Even if the strange vocabulary learning test had some predictive power, it would be a poor assessment when compared to more comprehensive analyses of the full range of underlying cognitive processes (e.g. reading complex and relevant passages in one's native language). Ultimately, statistical numeracy tests are robust predictors

because their imposed task demands are more representative of the diverse sub-skills and processes typically involved in effective naturalistic decision-making (e.g. practical and personally meaningful inductive reasoning and self-regulation under risk and uncertainty). For these and many other reasons, in the light of emerging data it is likely that more comprehensive psychometric studies will necessitate major restructuring of unified models of general human intelligence and the nature of its constituent abilities, as depicted in Figure 26.5. This figure depicts a hypothetical restructuring of Carroll's (1993) cognitive ability model that is predicted to emerge as statistical numeracy and general decision-making skill are more broadly and accurately represented in factor-analytic studies of general intelligence. As depicted, based on the extant evidence, the structure indicates that Carroll's current *Fluid Intelligence* factor appears likely to dissolve by dividing its factor loadings between acquired *General Decision-Making Skill* (e.g. practical logical inductive reasoning skills) and *General Fluid Memory and Learning* (e.g. executive functioning, attentional control), consistent with recent studies on the relations between decision-making skill, fluid intelligence, and numeracy. The model would also better accord with research indicating that the tests commonly used to estimate fluid intelligence (e.g. matrix reasoning tasks) tend to largely be measures of variations in coordinated general attentional control and short-term (working) memory storage (McCabe et al., 2010). To further explore the roles of skills and abilities in representative decision-making tasks, we now turn to recent research on risk communication, training, and decision support.

To further explore the roles of skills and abilities in representative decision-making tasks, we turn to recent research on risk communication, training, and decision support.

Simple, Powerful Decision Support

Numerate decision-makers are generally more resilient against information distortion and choice manipulation effects. Unfortunately, a large proportion of highly educated and intelligent working professionals, such as physicians, have relatively low levels of statistical numeracy. This weakness translates into misinterpretations of probability expressions, causing misunderstandings and potentially dangerous risk communication biases (Garcia-Retamero & Cokely, 2017; Garcia-Retamero et al., 2016b). For example, research indicates that nearly 50 percent of sampled physicians cannot correctly answer questions like "if person A's risk of getting a disease is 1 percent in ten years, and person B's risk is double that of A's, what is B's risk?" Nevertheless, an impressive number of studies using simple graphical representations of numerical expressions of probability (bar and line charts, and icon arrays) have been found to improve decision-making quality in diverse professionals, patients, and publics (for reviews see Garcia-Retamero & Cokely, 2013, 2017). In one influential study involving a large nationally representative sample from the United States, results indicated that providing visual aids in addition to numerical information about the effectiveness of medical treatments increased the accuracy of less numerate people's judgments from less than 20 percent to nearly 80 percent. The benefits of visual aids essentially equated the more and less numerate individuals, as long as participants had some minimal graph interpretation skills (i.e. graph literacy). In accord with Skilled Decision Theory, visual aids and other

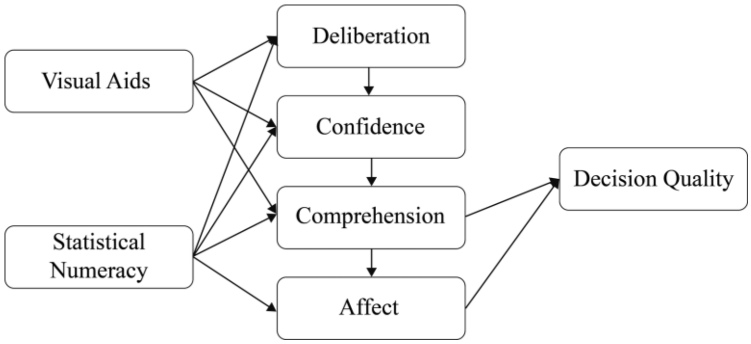

Figure 26.6 A generalized structural process model of skilled decision-making, depicting typical patterns of situational, cognitive, and emotional interdependencies in accord with Skilled Decision Theory.

transparent decision aids will tend to causally eliminate large differences in skilled decision-making by increasing deliberative evaluation and representative understanding of risks in long-term memory(Garcia-Retamero, Cokely, & Hoffrage, 2015). For example, in a recent study of some 300 practicing orthopedic surgeons, deliberation was found to partially mediate the relations between numeracy and clinically relevant judgment in a control condition (i.e. interpretation of actual anesthesia risks based on a recent peer-reviewed clinical trial). However, in a separate condition, differences in decision performance were completely eliminated when people were given a simple visual aid that promoted better understanding and more extensive deliberation among less skilled surgeons.

Transparent decision aids are specifically designed to generate an accurate, robust, and *representative understanding* of risks and information in long-term memory (e.g. promote a thoroughly integrated and balanced understanding of relevant trade-offs, options, data, and consequences). Thus, visual aids promote risk literacy (i.e. the ability to evaluate and understand risk) in many of the same ways statistical numeracy skills do, as can be seen in the structural process model depicted in Figure 26.6. The evidence showing that transparent decision aids causally improve decision-making by enhancing representative understanding and domain-specific risk literacy, such as health risk literacy and natural hazard risk literacy, comes from many different studies (for reviews see Garcia-Retamero & Cokely, 2013, 2014, 2017). All studies were explicitly designed to be naturalistic and ecologically valid, accurately reproducing actual problems that people commonly encounter when they evaluate personally relevant information about health, money, relationships, and the like. For example, tasks have included investigations on the accuracy of perceptions of health and disease risk, and risk reductions; inferences about the predictive power of medical tests and treatment effectiveness; assessments of subjective confidence in choices and risk perceptions; evaluations and integrations of emotions, trust, and information accuracy; assessment and trajectories of health outcomes; preferences and memory for health information; and changes in attitudes, behavioral intentions, behaviors, and high-stakes informed decisions.

Figure 26.6 depicts a structural process model of skilled decision-making derived from causal experimentation and formal quantitative structural equation modeling of high-stakes decision-making in health, public policy, and professional domains (e.g. medical screening, cancer treatment, responses to the Ebola pandemic). Numeracy and decision aids (e.g. visual aids) support skilled decision-making both directly and indirectly through metacognitive and risk comprehension effects (e.g. deliberation, confidence monitoring, and calibration), refining representative understanding and affective reactions (Garcia-Retamero et al., 2015). Numeracy also exerts a direct effect on affective responses such that both risk comprehension and numeracy independently influence the precision and calibration of affective intuitions (e.g. the relative subjective evaluation of positive and negative emotional reactions). In accord with *Skilled Decision Theory*, the model indicates that high-quality skilled and informed decision-making primarily follows from the interplay of deliberative evaluation (e.g. cognitive and metacognitive elaboration and exploration) and representative understanding (e.g. precise and accurate integrated mental models of risks that inform metacognitive and affective deliberation), providing essential foundations for adaptive heuristic decision making (e.g. a representative understanding allows one to truly feel what's at stake, and why, thereby ordering and integrating exploration, cues, and priorities so as to calibrate fast and frugal decision-making heuristics that rely on non-exhaustive lexicographic search; Gigerenzer & Gaissmaier, 2011).

Efficient General Skill Training

As an extension of our work on the benefits of visual aids we have recently developed an online tutoring program that helps people learn how to interpret and use various graphs like bar charts, line charts, icon arrays, and decision trees – i.e. a graph literacy training system (Cokely et al., forthcoming; Woller-Carter, 2016; Ybarra et al., 2017). In theory, by developing an efficient and relatively brief (less than two hours) graph literacy training program accessible to a wide range of individuals, we reasoned that we might be able to reach and empower more people. That is, we aimed to help improve graph literacy directly so that more people could benefit from simple visual aids that are highly effective for many types of risk communications, including those who score lower on standardized general ability and intelligence tests.

Beyond the many practical benefits, a noteworthy theoretical result of our graph literacy training program was that training translated directly into large improvements in specific types of skilled decision-making. For example, in one experiment about 100 participants completed either the graph literacy training program or a study skills training program. They were then tested on ostensibly unrelated decision-making tasks that did not include any type of visual aid. Participants in the graph literacy training condition showed very large improvements in scores on tests of framing effects, sunk costs, and class inclusion illusions as compared to the control condition (e.g. about 1.5 standard deviation). Structural analyses confirmed that the improvements we observed in skilled decision-making were mediated by improvements in graph literacy independent of other basic cognitive abilities (i.e. after statistically controlling for numeracy, fluid intelligence, and other cognitive abilities) (Cokely et al., forthcoming; Woller-Carter, 2016). Theoretically, risk literacy is the central necessary and potentially sufficient condition for skilled and informed decision-making among healthy and motivated adults in naturalistic general decision settings. The fact that graph literacy training substantially improved aspects of general decision-making skill may seem surprising (Simons et al., 2016). However, we anticipated these effects based on the findings showing that visual aids tend to be particularly beneficial for these types of tasks (e.g. bar charts de-bias framing effects, icon arrays de-bias ratio biases, decision trees de-bias sunk cost effects). In some real sense when someone learns how to use graphs to represent data they are learning how to build useful mental models of risks and relations in their long-term memory. Beyond the benefits of graph literacy training programs, there are many effective programs designed to directly develop essential probabilistic reasoning skills (Paas, 1992; Rittle-Johnson & Koedinger, 2001). Although it can be difficult, there is no doubt that skilled decision-making can be dramatically improved with the right kinds of guidance, motivation, and deliberate practice (Arkes, 1991; Baron & Brown, 2012; Brase, 2014; Chang, Chen, Mellers, & Tetlock, 2016; Clegg et al., 2015; Eskreis-Winkler et al., 2016; Fong, Krantz, & Nisbett, 1986; Garcia-Retamero, Cokely, Ghazal & Joeris, 2016a; Larrick, 2014; Larrick, Morgan, & Nisbett, 1990; Mellers et al. 2014, 2015a, 2015b; Morewedge et al., 2015; Peters, 2017; Peters et al., 2017; Soll, Milkman, & Payne, 2015; Torgerson, Porthouse, & Brooks, 2005; Xin & Jitendra, 1999).

Conclusions

For more than a century people have used theoretical assumptions to argue that general intelligence constrains decision-making quality, causing substantial differences in human potential and outcomes. In turn, some have argued that such associations should partially dictate the structure of our policies, rights, institutions, and welfare practices. Many scientists have endeavored to map underlying issues in transparent ways. Many others have used these assumptions as justification for racist, sexist, and violent discrimination, disenfranchising millions of people and minority groups (Gould, 1996; Nisbett et al., 2012). Setting aside moral and ethical outrage, at the heart of the scientific issue is a basic question about the extent to which abilities like fluid intelligence *actually* constrain decision-making quality.

While theory of the past was built on broad assumptions about observed correlations, emerging experiments, training programs, and cognitive process tracing studies provide converging causal evidence. Contrary to long-held assumptions, skilled decision-making does not generally require high levels of fluid intelligence or special abstract reasoning capacities. At an extreme, we find overwhelming evidence in decades of research on verifiable expert performers. If there are any effects of general cognitive abilities on expert decision-makers, they are increasingly trivialized and hard to detect compared to the profound decision quality benefits that accumulate with deliberate practice and the acquisition of specialized skills and knowledge (Ericsson et al., 2006). Even among non-experts, evidence indicates that decision-making is an acquired skill that generally operates independent of fluid intelligence. One's acquired level of statistical numeracy in particular, as measured by the Berlin Numeracy Tests and others, tends to be the single strongest predictor of general decision-making skill across laboratory, naturalistic, and real-world contexts. This finding is consistent with the extensive evidence showing that quantitative skills are among the most influential educational variables associated with advancing economic prosperity in industrialized countries (Hanushek & Woessmann, 2010; Hunt & Wittmann, 2008). Results also reveal that simple interventions and brief training programs (e.g. visual aids and adaptive computerized tutors) can dramatically improve skilled decision- making of diverse individuals who vary widely in abilities, proficiencies, educations, back- grounds, values, and countries of residence (Garcia-Retamero & Cokely, 2011, 2013, 2017; see also Bruine de Bruin & Bostrom, 2013; Bruine de Bruin et al., 2007; Fischhoff, 2013; Fischhoff, Brewer, & Downs, 2012; Peters, 2017; Petrova et al., 2014, 2015; Trevena et al., 2013).

It is noteworthy that statistical numeracy and other practical inductive reasoning tasks have historically been neglected in psychometric intelligence research, given that these are the primary factors linking quantitative abilities and general decision-making skill. This robust association in part reflects shared elements on surface levels (e.g. numbers are common features of many decisions) and at foundations (e.g. understanding risk and uncertainty). Consistent with Skilled Decision Theory, statistical numeracy predicts skilled decision-making because both require practical probabilistic reasoning and skilled metacognition (e.g. accurately evaluating and integrating thoughts, feelings, risks, and values). When these skills are combined with personally meaningful deliberation, they tend to promote a thorough and robustly representative understanding (e.g. adaptive mental models and integrated situation awareness), circumventing basic attentional capacity limitations with well-organized knowledge structures in long-term memory (e.g. vast personal knowledge and experience). Ultimately, a representative understanding of a decision problem helps people intuitively and precisely feel the relative weights of the essential issues, thereby enabling the adaptive use of heuristic evaluation and decision strategies.

Understanding Risks

Despite recent advances in our understanding of skilled decision-making, several obstacles present great challenges. Strongly polarized beliefs and values exist in most communities and countries. Consistent with the current review, improvements in communications and clarifications of material facts are likely to resolve many of these issues. Nevertheless, research also indicates that some of these biases may never go away (Feltz & Cokely, 2009, 2012, 2013, 2016). Some fundamental morally relevant biases are so robust they persist even among skilled reasoners, including verifiable experts who have devoted their lives to understanding the specific and relevant philosophical issues (Cokely & Feltz, 2009a, 2009b, 2014; Kahan, Jenkins-Smith, & Braman, 2011; Schulz, Cokely, & Feltz, 2011). In part, such biases persist because we don’t have reliable methods that can provide accurate feedback about the objective truth of many moral questions (if it even exists). We can precisely measure things like the weight of gold, but the same cannot be said for assessments of issues like rightness, justice, or equality. Even though skilled decision-makers can approximate normatively superior decision-making standards, these standards cannot tell us which decisions we should make unless we know what we should value. Moreover, decision sciences and technologies are rapidly enhancing the ability of some to control and manipulate our choices, often without our awareness or consent. While some valuable efforts are being deployed to help more of us make beneficial decisions without limiting our choices (Thaler & Sunstein, 2008), many others are designed to promote special interests and values via nonrational persuasion and choice architecture manipulation.

As decision science advances we will increasingly be faced with a fundamental ethical question: *Who should decide how we live our lives?* A straightforward answer is that every person who is competent should have the opportunity to make their own decisions, given basic conditions (e.g. absent unwanted infringement on the autonomy of others). This perspective accords with widely accepted standards for ethical and informed decision-making that emphasize autonomy and the importance of opportunities for deliberation in the light of one’s own values (Drane, 1984). If we aspire to satisfy these ethical standards and promote related democratic ideals, decision-makers will need a balanced but not necessarily extensive understanding of the risks and implications of various courses of action – i.e. *a representative understanding* (Feltz, 2015; Feltz & Cokely, 2016; Fischhoff et al., 2011; Garcia-Retamero & Cokely, 2017). Developing efficient scientific means of identifying and promoting representative understanding across evolving domains and conflicts is a formidable and worthy task. Methods developed for research on expert performance can help light the way for those decision scientists who dare to take on informed decision-making challenges in complex, controversial, and high-stakes domains (e.g. climate change, energy, education, health, cybersecurity). To the extent we improve representative understanding, the research reviewed here indicates that diverse people will be better prepared to make and discuss life-altering decisions in adaptive and ethical ways. Over time, even small improvements in skilled decision-making add up to substantial personal and societal benefits. It is a great gift and responsibility to know that nearly everyone has the ability to make well-informed and skilled decisions so long as they understand risks.

Author Note

Partial financial support for this research was provided by grants from the US National Science Foundation (SES-1253263), the US National Academies of Science, Medicine, & Engineering, the Templeton Foundation, and the Ministerio de Economía y Competitividad (PSI2011-22954 & PSI2014-51842-R), with substantial support from the Max Planck Institute for Human Development (Berlin, DE), and the National Institute for Risk & Resilience at the University of Oklahoma. We thank K. Anders Ericsson, Jon Baron, Robert Hoffman, and several anonymous reviewers, and other colleagues, for valuable feedback. We offer special thanks to Gloria Rennie, Kelvin Droegemeier, and Gerd Gigerenzer for pivotal support and vision. We thank Carol Silva, Jessica Pope, Jeane Hardy, Nina Carlson, and Hank Jenkins-Smith for valuable leadership, administrative, and technical contributions. Finally, we offer gratitude for the great efforts of all affiliated scientists, advisors, and community members who contribute to the international www.RiskLiteracy.org project. Correspondence concerning this chapter should be addressed to Dr. Edward T. Cokely, National Institute for Risk & Resilience, 201 Stephenson Parkway, 5PP Suite 2300, Norman, Oklahoma 73019. Contact: cokely@ou.edu.

References

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, 110, 486–498.

Arkes, H. R. (2016). A levels of processing interpretation of dual-system theories of judgment and decision making. *Theory & Psychology*, 26, 459–475.

Baron, J. (1985). *Rationality and intelligence*. Cambridge University Press.

Baron, J. (2008). *Thinking and deciding* (4th edn.). Cambridge University Press.

Baron, J., & Brown, R. V. (eds.) (2012). *Teaching decision making to adolescents*. New York: Routledge.

Brase, G. L. (2014). The power of representation and interpretation: Doubling statistical reasoning performance with icons and frequentist interpretations of ambiguous numbers. *Journal of Cognitive Psychology*, 26, 81–97.

Breakspear, S. (2012). *The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance*. OECD Education Working Paper No. 71. Paris: OECD Publishing.

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

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

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Chang, W., Chen, E., Mellers, B., & Tetlock, P. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment and Decision Making*, 11, 509–526.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Science*, 10, 287–291.

Clegg, B. A., McKernan, B., Martey, R. M., Taylor, S. M., Stromer-Galley, J., Kenski, K., ... & Shaw, A. (2015). Effective mitigation of anchoring bias, projection bias, and representativeness bias from serious game-based training. *Procedia Manufacturing*, 3, 1558–1565.

Cokely, E. T. (2009). Beyond generic dual processes: How should we evaluate scientific progress? *PsycCritiques*, 54, Article 10.

Cokely, E. T., Allan, J., Ghazal, S., Feltz, A., & Garcia-Retamero, R. (Forthcoming). General decision making skill and numeracy: Component abilities.

Cokely, E. T., & Feltz, A. (2009a). Adaptive variation in judgment and philosophical intuition. *Consciousness and Cognition*, 18, 356–358.

Cokely, E. T., & Feltz, A. (2009b). Individual differences, judgment biases, and theory-of-mind: Deconstructing the intentional action side effect asymmetry. *Journal of Research in Personality*, 43, 18–24.

Cokely, E. T., & Feltz, A. (2014). Expert intuition. In L. M. Osbeck & B. S. Held (eds.), *Rational intuition: Philosophical roots, scientific investigations* (pp. 213–238). Cambridge University Press.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, 7, 25–47.

Cokely, E. T., Ghazal, S., Galesic, M., Garcia- Retamero, R., & Schulz, E. (2013). How to measure risk comprehension in educated samples. In R. Garcia-Retamero & M. Galesic (eds.), *Transparent Communication of Health Risks: Overcoming Cultural Differences* (pp. 29–52). New York: Springer.

Cokely, E. T., Ghazal, S., & Garcia-Retamero, R. (2014). Measuring numeracy. In B. L. Anderson & J. Schulkin (eds.), *Numerical reasoning in judgments and decision making about health* (pp. 11–38). Cambridge University Press.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, 4, 20–33.

Cokely, E. T., Kelley, C. M., & Gilchrist, A. H. (2006). Sources of individual differences in working memory: Contributions of strategy to capacity. *Psychonomic Bulletin & Review*, 13, 991–997.

Cokely, E. T., Schooler, L. J., & Gigerenzer, G. (2010). Information use for decision making. In M. N. Maack & M. J. Bates (eds.), *Encyclopedia of Library and Information Sciences* (3rd edn.) (pp. 2727–2734). New York: Taylor & Francis.

Del Missier, F., Mäntylä, T., & Bruin, W. B. (2012). Decision-making competence, executive functioning, and general cognitive abilities. *Journal of Behavioral Decision Making*, 25, 331–351.

Drane, J. F. (1984). Competency to give an informed consent: A model for making clinical assessments. *Journal of the American Medical Association*, 252, 925–927.

Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16, 939–944.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, 51, 380–417.

Ericsson, K. A. (ed.) (1991). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Mahwah, NJ: Erlbaum.

Ericsson, K. A., Charness, N., Hoffman, R. R., & Feltovich, P. J. (eds.) (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, 208, 1181–1182.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard Business Review*, 85, 114–121.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215–251.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. Cambridge, MA: MIT Press.

Eskreis-Winkler, L., Shulman, E. P., Young, V., Tsukayama, E., Brunwasser, S. M., & Duckworth, A. L. (2016). Using wise interventions to motivate deliberate practice. *Journal of Personality and Social Psychology*, 111, 728–744.

Evans, J. S. B., & Frankish, K. E. (2009). *In two minds: Dual processes and beyond*. Oxford University Press.

Feltz, A. (2015). Ethical information transparency and sexually transmitted infections. *Current HIV Research*, 13, 421–431.

Feltz, A., & Cokely, E. T. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness and Cognition*, 18, 342–350.

Feltz, A., & Cokely, E. T. (2012). The philosophical personality argument. *Philosophical Studies*, 161, 227–246.

Feltz, A., & Cokely, E. T. (2013). Predicting philosophical disagreement. *Philosophy Compass*, 8, 978–989.

Feltz, A., & Cokely, E. T. (2016). Personality and philosophical bias. In J. Sytsma & W. Buckwalter (eds.), *A Companion to Experimental Philosophy* (pp. 578–589). New York: John Wiley.

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

Fischhoff, B., Brewer, N. T., & Downs, J. T. (2012). *Communicating risks and benefits: An evidence-based user's guide*. Silver Spring, MD: US Food and Drug Administration.

Fisher, R. A. (1952). *The design of experiments*. London: Macmillan.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, 18, 253–292.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, 137, 316–344.

Fox, M. C., & Mitchum, A. L. (2013). A knowledge-based theory of rising scores on "culture-free" tests. *Journal of Experimental Psychology: General*, 142, 979–1000.

Fox, M. C., & Mitchum, A. L. (2014). Confirming the cognition of rising scores: Fox and Mitchum (2013) predicts violations of measurement invariance in series completion between age-matched cohorts. *PloS One*, 9, e95780.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19, 25–42.

Gal, I. (2003). Teaching for statistical literacy and services of statistics agencies. *American Statistician*, 57, 80–84.

Garcia-Retamero, R., & Cokely, E. T. (2011). Effective communication of risks to young adults: Using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology: Applied*, 17, 270–287.

Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science*, 22, 392–399.

Garcia-Retamero, R., & Cokely, E. T. (2014). The influence of skills, message frame, and visual aids on prevention of sexually transmitted diseases. *Journal of Behavioral Decision Making*, 27, 179–189.

Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: A systematic review of health research and evidence-based design heuristics. *Human Factors*, 59, 582–627.

Garcia-Retamero, R., Cokely, E. T., Ghazal, S., & Joeris, A. (2016a). Measuring graph literacy without a test: A brief subjective assessment. *Medical Decision Making*, 36, 854–867.

Garcia-Retamero, R., Cokely, E. T., & Hoffrage, U. (2015). Visual aids improve diagnostic inferences and metacognitive judgment calibration. *Frontiers in Psychology*, 6, 932.

Garcia-Retamero, R., Cokely, E. T., Wicki, B., & Joeris, A. (2016b). Improving risk literacy in surgeons. *Patient Education and Counseling*, 99, 1156–1161.

Garcia-Retamero, R., Wicki, B., Cokely, E. T., & Hanson, B. (2014). Factors predicting surgeons' preferred and actual roles in interactions with their patients. *Health Psychology*, 33, 920–928.

Ghazal, S. (2014). Component numeracy skills and decision making. PhD Dissertation, Michigan Technological University.

Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making*, 9, 15–34.

Gigerenzer, G. (2015). *Risk savvy: How to make good decisions*. New York: Penguin.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–482.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. Oxford University Press.

Ginsburg, L., Manly, M., & Schmitt, M. J. (2006).
The components of numeracy. NCSALL Occasional Paper. Boston, MA: National Center for the Study of Adult Learning and Literacy.

Goldman, S. R., & Pellegrino, J. W. (1984). Deductions about induction. *Advances in the Psychology of Human Intelligence*, 2, 149–197

Gould, S. J. (1996). *The mismeasure of man*. New York:
W. W. Norton.

Hacking, I. (2006). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge University Press.

Hanushek, E. A., & Woessmann, L. (2010). *The economics of international differences in educational achievement*. NBER Working Paper No. 15949. Cambridge, MA: National Bureau of Economic Research.

Hastie, R., & Dawes, R. M. (eds.) (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Thousand Oaks, CA: Sage Publications.

Heckman, J. J. (1995). Lessons from the bell curve. *Journal of Political Economy*, 103, 1091−1120.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Holyoak, K. J., & Morrison, R. G. (eds.) (2005). *The Cambridge handbook of thinking and reasoning*. Cambridge University Press.

Huff, D. (1954). *How to lie with statistics*. New York: W. W. Norton.

Hunt, E., & Wittmann, W. (2008). National intelligence and national prosperity. *Intelligence*, 36, 1−9.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences USA*, 108, 10081−10086.

Jasper, J. D., Bhattacharya, C., & Corser, R. (2017). Numeracy predicts more effortful and elaborative search strategies in a complex risky choice context: A process-tracing approach. *Journal of Behavioral Decision Making*, 30, 224−235.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press.

Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14, 147−174.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697−720.

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

Keller, N., Cokely, E. T., Katsikopoulos, K. V., & Wegwarth, O. (2010). Naturalistic heuristics for decision making. *Journal of Cognitive Engineering and Decision Making*, 4, 256−274.

Klein, G. (1999). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Kutner, M., Greenburg, E., Jin, Y., & Paulsen, C. (2006). The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy. NCES 2006-483. Washington, DC: National Center for Education Statistics.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389−433.

Larrick, R. P. (2014). Debiasing. In D. J. Koehler & N. Harvey (eds.), *Blackwell handbook of judgment and decision making* (pp. 316−338). Malden, MA: Blackwell Publishing.

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science*, 1,362−370.

Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science*, 26, 1827−1832.

Lindskog, M., Winman, A., Juslin, P., & Poom, L. (2013). Measuring acuity of the approximate number system reliably and validly: The evaluation of an adaptive test procedure. *Frontiers in Psychology*, 4, Article 510.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21, 37−44.

McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24, 222−243.

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., ... & Tetlock, P. (2015a). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied*, 21, 1−14.

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., .. . & Ungar, L. (2015b). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science*, 10, 267−281.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., & ... Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, 25, 1106−1115.

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.

Moshman, D. (2000). Diversity in reasoning and rationality: Metacognitive and developmental considerations. *Educational Psychology Papers and Publications*, 46, 689−690.

Moxley, J. H., Ericsson, K. A., Charness, N., & Krampe, R. T. (2012). The role of intuition and deliberative thinking in experts' superior tactical decision-making. *Cognition*, 124, 72−78.

Newall, P. W. (2016). Downside financial risk is misunderstood. *Judgment and Decision Making*, 11, 416−422.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, 67, 130−159.

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2015). Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. *Journal of Experimental Psychology: Applied*, 21, 178−194.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11, 988−1010.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84, 429−434.

Pachur, T., & Galesic, M. (2013). Strategy selection in risky choice: The impact of numeracy, affect, and cross-cultural differences. *Journal of Behavioral Decision Making*, 26, 260−271.

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

Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences*. London: Macmillan.

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, 21, 31−35.

Peters, E. (2017). Educating good decisions. *Behavioural Public Policy*, 1, 162−176.

Peters, E., & Bjalkebring, P. (2015). Multiple numeric competencies: When a number is not just a number. *Journal of Personality and Social Psychology*, 108, 802−822.

Peters, E., Hibbard, J., Slovic, P., & Dieckmann, N. (2007). Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs*, 26, 741−748.

Peters, E., Shoots-Reinhard, B., Tompkins, M. K., Schley, D., Meilleur, L., Sinayev, A., … & Crocker, J. (2017). Improving numeracy through values affirmation enhances decision and STEM outcomes. *PloS ONE*, 12, e0180674.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17, 407−413.

Petrova, D., Garcia-Retamero, R., Catena, A., Cokely, E. T., Herredia Carrasco, A., Arrebola Moreno, A., & Ramírez Hernández, J. A. (2017a). Numeracy predicts risk of pre-hospital decision delay: A retrospective study of acute coronary syndrome survival. *Annals of Behavioral Medicine*, 51, 292−306.

Petrova, D., Garcia-Retamero, R., & Cokely, E. T. (2015). Understanding the harms and benefits of cancer screening: A model of factors that shape informed decision making. *Medical Decision Making*, 35, 847−858.

Petrova, D., Kostopoulou, O., Delaney, B. C., Cokely, E. T., & Garcia-Retamero, R. (2017b). Strengths and gaps in physicians' risk communication: A scenario study of the influence of numeracy on cancer screening communication. *Medical Decision Making*. DOI: 0272989X17729359.

Petrova, D., van der Pligt, J., & Garcia- Retamero, R. (2014). Feeling the numbers: On the interplay between risk, affect, and numeracy. *Journal of Behavioral Decision Making*, 27, 191−199.

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty*. New York: Random House

Reyna, V. F. (2004). How people make decisions that involve risk: A dual process approach. *Current Directions in Psychological Science*, 13, 60−66.

Reyna, V. F. (2008). A theory of medical decision making and health: Fuzzy trace theory. *Medical Decision Making*, 28, 850−865.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, 7, 1−75.

Reyna, V. F., Nelson, W. L., Han P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, 135, 943−973.

Rittle-Johnson, B., & Koedinger, K. R. (2001). Using cognitive models to guide instructional design: The case of fraction division. *Proceedings of the Cognitive Science Society*, 23.

Rubinstein, A. (2013). Response time and decision making: An experimental study. *Judgment and Decision Making*, 8, 540−551.

Savage, L. J. (1954). *The foundations of statistics*. New York: John Wiley.

Schlaifer, R., & Raiffa, H. (1961). *Applied statistical decision theory*. Cambridge, MA: Harvard University Press.

Schley, D. R., & Peters, E. (2014). Assessing "economic value": Symbolic-number mappings predict risky and riskless valuations. *Psychological Science*, 25, 753−761.

Schulz, E., Cokely, E. T., & Feltz, A. (2011). Persistent bias in expert judgments about free will and moral responsibility: A test of the expertise defense. *Consciousness and Cognition*, 20, 1722–1731.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127, 966–972.

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

Simon, H. A. (1957). A bounded-rationality model of rational choice. In H. A. Simon, *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: John Wiley.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, 41, 1–20.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, 17, 103–186.

Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A user's guide to debiasing. In G. Keren & G. Wu (eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 924–951). Malden, MA: Wiley-Blackwell.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. New York: Psychology Press.

Stanovich, K. E. (2016). The comprehensive assessment of rational thinking. *Educational Psychologist*, 51, 23–34.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and Brain Sciences*, 23, 701–717.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: Toward a test of rational thinking*. Cambridge, MA: MIT Press.

Steen, L. A. (ed.) (1990). *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academies Press.

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

Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, 63, 107–140.

Torgerson, C., Porthouse, J., & Brooks, G. (2005). A systematic review of controlled trials evaluating interventions in adult literacy and numeracy. *Journal of Research in Reading*, 28, 87–107.

Traczyk,J., & Fulawka, K. (2016). Numeracy moderates the influence of task-irrelevant affect on probability weighting. *Cognition*, 151, 37–41.

Trevena, L. J., Zikmund-Fisher, B. J., Edwards, A., Gaissmaier, W., Galesic, M., Han, P. K., .. . & Ozanne, E. (2013). Presenting quantitative information about decision outcomes: A risk communication primer for patient decision aid developers. *BMC Medical Informatics and Decision Making*, 13, S2–S7.

Tversky, A., & Kahneman, D. (1985). The framing of decisions and the psychology of choice. In V. T. Covello, J. L. Mumpower, P. J. M. Stallen, & V. R. R. Uppuluri (eds.), *Environmental impact assessment, technology assessment, and risk analysis* (pp. 107–129). Berlin: Springer.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, 26(2), 198–212.

Woller-Carter, M. (2016). Development of the intelligent graphs for everyday risky decisions tutor. PhD Dissertation, Michigan Technological University.

Woller-Carter, M., Okan, Y., Cokely, E. T., & Garcia-Retamero, R. (2012). Communicating and distorting risks with graphs: An eye-tracking study. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 1723–1727). Newbury Park, CA: Sage Publications.

Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *Journal of Special Education*, 32, 207–225.

Yates, J. F. (1990). *Judgment and decision making*. Upper Saddle River, NJ: Prentice-Hall.

Ybarra, V. T., Cokely, E. T., Adams, C., Woller-Carter, M., Allan, J. N., Feltz, A., & Garcia-Retamero, R. (2017). Training graph literacy: Developing the RiskLiteracy.org Outreach Platform. *Proceedings of the Cognitive Science Society*, 29.