24

Expertise in Decision Making

Richard P. Larrick
Fuqua School of Business, Duke University, USA

Daniel C. Feiler
Tuck School of Business, Dartmouth College, USA

Introduction

People face many consequential judgments and decisions for which they feel they lack the requisite knowledge. How should I invest my retirement savings? Why is my car making that noise? Will it rain during the family reunion on Saturday? Why does my daughter have a fever? For many such cases, people turn to someone else – an expert whom they trust to reach a better answer than they can reach themselves. But to what degree is trust warranted? This chapter reviews the literatures on expertise and on decision making to consider the nature and development of decision making expertise, including its strengths and weaknesses.

Research on judgment and decision making (JDM) often portrays a pessimistic view of decision making ability: decisions are the product of cognitive shortcuts that can produce systematic and consequential errors (Kahneman, 2003). In turn, this pessimistic view has occasionally provoked a more optimistic rebuttal – ordinary cognition is well adapted to process limited information quickly and accurately (Gigerenzer, 2007). A similar debate has emerged in the literature on expert decision making (Kahneman & Klein, 2009). Researchers who have grounded their perspectives in JDM research have taken a more skeptical view of expert judgment (Camerer & Johnson, 1991; Tetlock, 2005). Other perspectives, such as the naturalistic decision making (NDM) approach of Gary Klein and his colleagues (Klein, 1998), have offered a positive account of the abilities of experts. Despite the apparent differences, a clear consensus has emerged since the 1990s (Hogarth, 2001; Kahneman & Klein, 2009; Shanteau, 1992) – the key issue is not whether expertise in decision making exists but that it emerges only under specifiable conditions. The main goal of this chapter is to provide a framework for identifying when expertise in decision making can emerge. Such an understanding is practically useful for seeking expertise in others or for striving to build it in oneself.

The Wiley Blackwell Handbook of Judgment and Decision Making, First Edition.
Edited by Gideon Keren and George Wu.

Defining Expertise

Following Ericsson (2006), we define expertise as the possession of domain-specific knowledge that is acquired through experience or training and that leads to superior, reproducible performance in domain-related tasks. This definition has three key elements that we build on in the rest of the chapter. First, expertise is *domain specific*. Research on expertise has focused on a wide range of activities, including accounting, chess, fire-fighting, Go, medicine, software programming, tennis, and typing. In each of these domains, experts differ from novices in that they have a sophisticated understanding of their task and can quickly recognize important cues and sort through what is strategically or causally relevant to their decisions. However, this knowledge is limited to that specific domain. One of the main findings in the expertise literature is that there is little transfer of superior performance in one's area of expertise to other domains. The ability to escape checkmate is unrelated to the ability to detect fraud. We review the domain-specific, schematic nature of expertise in the section "Research on Expertise: Expertise is Schematic" of the chapter and explore some of the limitations of such knowledge in the section "Shortcomings of Expertise".

Second, expert knowledge is *acquired from experience and training*. In this respect, expertise differs from intelligence, which is characterized as pure reasoning ability. For expertise to arise from experience and training, decision makers must be exposed to experiences that provide immediate, accurate feedback about relationships in the world (Brehmer, 1980; Einhorn & Hogarth, 1978; Hogarth, 2001). Experience can provide accurate feedback only if the task is itself predictable (Hammond, 1996; Kahneman & Klein, 2009; Shanteau, 1992). These are essential themes we develop in the third section in this chapter: The Role of the Environment.

Third, expertise leads to *superior, reproducible performance*. This last element is not universally included in the definition or study of expertise (Ericsson, 2006). Some approaches to studying expertise rely on peer nominations of expertise (e.g., based on credentials or reputation) and focus on differences in the way experts and novices reason in a given situation. The danger with focusing on a type of reasoning as the standard for expertise is that an individual might be impeccably coherent and well-reasoned in a way that has no link to performance. For example, an "expert" astrologer can consistently (and perhaps eloquently) reason about the relationship between the stars and human behavior with no correspondence between their predictions and reality. Ericsson (2006), however, has argued for the inclusion of superior, reproducible performance as a key element of expertise. The *reproducibility* of superior performance is a key part of the standard because outcomes almost always have components of luck, especially in small samples, and true ability exists only if it is sustainable.

Some definitions of expertise focus on its social basis, inferred from reputation and confidence (Shanteau, 1992). Although these characteristics may be correlated with the other elements in the definition we are using, and may be useful in helping an expert be influential with others, they are not essential elements of expertise. In fact, one of the great challenges of using experts is the modest correlation between peoples' assertions of expertise and their subsequent performance (see Using Expertise, the sixth section of this chapter).

Research on Expertise: Expertise is Schematic

Modern research on expertise starts with DeGroot's famous study of chess masters (Chase & Simon, 1973; DeGroot, 1965). If one pictures a decision tree of all the possible moves and countermoves in chess, there is an exponentially explosive number of branches that count beyond billions. Do chess masters excel because they can calculate more rapidly than their opponents and search more deeply for the best move? DeGroot's research answered the question with a resounding No. DeGroot had chess masters "think aloud" while considering their next move in a realistic game setting. One of his key findings was that chess masters rapidly generated a few very good moves while novices actually considered a larger set of possible moves, although with less direction towards the most optimal few.

Chase and Simon (1973) argued that, through experience, chess masters had come to recognize a large number of game situations and the best responses to those situations. In support of this argument, Chase and Simon showed that chess masters could briefly view a chess board and then recall it more accurately than novices – only, however, if the board configuration was one that occurred in real games. Chess masters showed no enhanced ability for random board configurations. Chase and Simon argued that chess masters "chunked" meaningful configurations into larger wholes made up of related parts – in effect, the board had meaning in that it captured a moment in the strategic interaction that was both familiar and logical. The random board was essentially meaningless. The ability to encode information happens rapidly and is not harmed by subsequent cognitive demands, suggesting that it is stored in long-term memory (Charness, 1976). Subsequent research (Gobet & Simon, 1996) proposed that chess masters organize chunks of information into higher level schemas, or templates, that they use to encode the whole board.

Klein (1993, 1998) has referred to this combination of recognizing situations and rapidly generating good solutions as recognition-primed decision making (RPD). This general element of expertise has been found in many domains. For example, expert physicians are more likely than novices to reason "forward" from observed symptoms to the best explanation that covers them; by contrast, novices often start with a hypothesized cause and then reason "backward" to whether the symptoms fit it or not, which is a process that often produces a poorer fit between symptoms and diagnosis (Patel & Groen, 1986).

Although superior task memory is a signature of expertise, expert performance is not simply a product of memorization – a kind of grand parlor trick. Chess masters do not beat lesser chess players because they can quickly memorize board positions. The fact that they can quickly memorize board positions is a product of their superior understanding of the flow of countless games – they recognize a snapshot of a single moment in a sequence of related, key strategic decisions. Superior memory is a concomitant effect of what really distinguishes experts, which is possession of more sophisticated knowledge structures or schemas than novices. These underlying knowledge structures go beyond simple lists of facts to an understanding of the causal structure and causal dynamics of a domain (Feltovich, Prietula, & Ericsson, 2006).

To illustrate the idea of a knowledge structure, quickly read the following passage (Bransford & Johnson, 1972):

> The procedure is actually quite simple. First you arrange things into different groups. Of course one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities that is the next step, otherwise you are pretty well set. It is important not to overdo things. That is, it is better to do too few things at one time than too many. In the short run this may not seem important but complications can easily arise. A mistake can be expensive as well. At first the whole procedure will seem complicated. Soon, however, it will become just another facet of life. It is difficult to foresee any end to the necessity for this task in the immediate future, but then one can never tell. After the procedure is completed one arranges the materials into different groups again. Then they can be put into their appropriate places. Eventually they will be used once more and the whole cycle will then have to be repeated. However, that is a part of life.

How much of the passage can you recall? Can you turn away from this chapter and write it down in order and in full? (You may want to try.) The challenge is that this passage is written rather abstractly. It describes a number of steps, but they seem unrelated to each other. Most people have only modest success. There is no framework available for organizing the steps into a meaningful, causal flow. However, it turns out you are in fact an expert in this activity – you have a schema ready to use. The schema is "doing laundry." If you re-read the passage now, all of the seemingly random pieces cohere into a single, familiar activity with a set of goals and an underlying logic. When people are told the activity in advance of reading the paragraph, they do a good job of remembering the key steps and remembering them in order.

A schema is a cognitive structure that links different pieces of information into a set of meaningful relationships, especially cause-effect and goal-directed relationships. Experts possess a rich understanding of these relationships and use them to guide their thinking. For example, expert firefighters think about fires in terms of the events that produced them and the events that might follow, whereas novices think more about the features of the fire (such as color and heat) (Klein, 1998). Baseball experts have better memory than novices for the sequence of actions described in a real baseball event but show no enhanced memory for random sequences of baseball actions or for events unrelated to baseball (Chiesi, Spilich, & Voss, 1979). Expert nephrologists rely on their knowledge of basic science when they encounter a complex pattern of symptoms more than medical students do (Norman, Trott, Brooks, & Smith, 1994). And experts in physics reason through physics problems by thinking about the underlying principles that apply to the solution (e.g., conservation of momentum), whereas novices look at the superficial features of a problem (e.g., the problem involves an inclined plane) (Chi, Feltovich, & Glaser, 1981).

There is not a clear consensus view on whether experts use more cues than novices (Shanteau & Stewart, 1992). Think-aloud protocols typically reveal that, compared to novices, experts do consider more cues, better cues, see a deeper problem structure, and give a more coherent explanation (Feltovich et al., 2006). This is not overly surprising given that novices have less ammunition to draw on when they reason about a domain-specific problem. But one of the distinguishing features of expertise is that it often leads to rapid decisions with seemingly little time for reflection. The contemporary view is that experts possess superior long-term working memory (Ericsson & Kintsch, 1995) in their domain of expertise. This guides the selection of the most relevant information and allows for rapid storage and retrieval that is not constrained in the usual ways that limit short-term memory (Charness, 1976). A cognitive schema allows an expert to recognize situations and swiftly attend to the most relevant cues for that circumstance, facilitating both faster performance and better performance. Selective cue use is a hallmark of expertise (Feltovich et al., 2006).

A medical study provides an interesting illustration of rapid, selective cue use. Kulatunga-Moruzi, Brooks, and Norman (2004) presented physicians who possessed varying levels of expertise with dermatological symptoms in different formats for both typical and atypical diseases. They systematically varied whether the physicians saw a photograph of a dermatological ailment, a list of all of the cues present in the photograph (all of which were veridical but only some of which were relevant to the correct diagnosis), or the list of cues followed by the photograph. Residents (a novice group) gave more accurate diagnoses when they read the list of cues before seeing the photographs. Presumably the list ensured that they attended to features they might have overlooked when collecting their own information. However, family practitioners and dermatologists (the expert groups) were less accurate when using the list of cues (either alone or prior to looking at the photograph). The more expert judges fared better when they acquired information from a visual examination in their usual manner – suggesting that they had a routine that led them to attend appropriately to relevant cues and to ignore irrelevant cues.

Where do these knowledge structures come from? One of the central findings in the expertise literature is that domain-specific, reproducible, superior performance often requires extensive experience in a domain: approximately 10 years of experience, or 10,000 hours (Ericsson, 2006). This number has been observed in many domains, including chess, music, and medicine, but there is certainly variability across experts. For example, some chess experts have achieved grandmaster level in only 3,000 hours (Campitelli & Gobet, 2011). It is also important to note that this number has been arrived at by identifying experts and then calculating the hours that they spent participating in their domain. Since often only those who demonstrate natural ability will feel encouraged to proceed with training, the 10,000 hours could be thought of as the typical time needed for a gifted individual to become a master, as opposed to the median person.[1] However, by any account, the number of hours of experience needed to achieve expertise appears to be in the thousands.

The principle of needing thousands of hours of experience has been offered as a necessary but not sufficient condition for the development of the highest levels of expertise. The problem is that people "level off" in their performance improvement if their thousands of hours are not continually challenging them (Ericsson, 2004, 2006). For example, spending 10,000 hours seeing elementary medical cases, playing tennis beginners, or strumming simple pop music will not stretch one's capability to more masterful levels. Thus, both the amount and the challenge of the experience matter. Ericsson has referred to the need for "deliberate practice" as a key to achieving the highest levels of expertise. Deliberate practice has two key components: (a) it is initially done with awareness and intention, and (b) it is focused on a specific task that is designed to be different from what is familiar but is relevant to future challenges. For example, this might entail deliberately studying the symptoms and treatment for a new disease, practicing a new shot in tennis, or practicing new chords in music. In sum, repetition under identical circumstances provides improvement but there are diminishing returns to such experience; eventually variation in the content of experience and greater challenges within the relevant domain are needed to develop expertise. We will go into greater depth on learning environments and training in the section on the role of the environment.

The process of acquiring thousands of hours of practice is familiar to everyone. Most readers have acquired some level of expertise in a life domain, such as math, writing, typing, driving, or a specific sport. There is a standard course to the learning that occurs in these tasks. Initial learning often involves a great deal of explicit instruction with many declarative statements about what to do ("hold your elbow up," or "check your blind spot before changing lanes"). These instructions are cognitively demanding to attend to. Because much of the information processing is slow and conscious, performance is impaired under cognitive load. Load itself is often increased for novices as they attend to situational features, concentrate on remembering key declarative statements, and are distracted by stress. But with repetition, people rely less on telling themselves a set of steps to follow, and the steps become automatic and habitual – they are easily executed without thought. Compared to teenagers in Driver's Ed, anyone with thousands of hours of actual driving experience is reasonably competent at checking blindspots, anticipating speed changes, and adjusting to weather conditions without deliberation. (Note that expert city drivers are not expert enough to take on pro formula racing because they do not have 10,000 hours spent in such conditions. This is an example of "leveling off" – it would take substantial deliberate practice under more difficult conditions to get to the Indy 500.)

This transition from deliberate learning to intuitive response puts this chapter of the handbook in interesting contrast with a common view in decision making research. One of the most central and useful frameworks in JDM has been the distinction between System 1 processes, which are intuitive, rapid, and automatic, and System 2 processes which are slower, effortful, and conscious (Kahneman, 2003, 2011; Sloman, 1996; Stanovich & West, 1998; see Keren & Schul, 2010, for a critical evaluation of the two systems framework). In most JDM research, processes grouped under System 2 are assumed to lead to superior decisions – those grouped under System 1, though efficient, are too easily influenced by irrelevant contextual factors, and System 2 must monitor and correct for System 1's mistakes. The development of expertise inverts these roles. Initial learning is highly dependent on System 2 processes (slow, effortful, conscious) and easily goes awry under cognitive load. With repetition, steps that had to be performed consciously become rapid, automatice System 1 processes.[2]

Thus, the essential feature of gaining thousands of hours of practice is a transition from *thinking* about elements in a task to intuitively and effectively *doing* them. This is what cognitive psychologists refer to as converting declarative knowledge to procedural knowledge. This conversion has interesting consequences for experts. One consequence is that inducing an expert to reflect on his or her behavior may impair performance. In his classic work on "gamesmanship," Stephen Potter (1947) proposed that one can undermine a sports opponent by calling attention to his or her technique. In golf or tennis, one can compliment an aspect of the opponent's swing – thereby leading the opponent to be overly aware of that element and break his or her normal flow. A second consequence of routinization is that experts struggle to describe what they are doing or how they make decisions. When researchers try to build models of expert judgment in artificial intelligence (AI) systems, they have to use multiple techniques to capture expertise because self-reports of general strategies are often incomplete or even inaccurate (Slovic, Fleissner, & Bauman, 1972). As a result, being expert in a task does not ensure that one can teach it. In fact, greater expertise moves one further from the perspective of the novice, which may even make one a *worse* teacher. As task ability becomes rapid and automatic, one loses awareness of how to decompose one's skills into specific steps. Teaching is itself a form of expertise and requires experience in how best to communicate strategies and knowledge to others.

In sum, research on expertise suggests that people can become expert in a specific task domain with a large amount of experience. Experience equips experts with schemas that direct their attention to relevant cues in that domain and allows them to rapidly diagnose problems, make forecasts, and choose effective actions in that domain. This summary of the expertise literature, however, begins to suggest some of the important limitations on expert decision making. Expertise depends on the quality of experience. The quality of experience is the critical moderator of the development of expertise, and we consider it at length in the next section. Expertise is schematic, and it inherits both encoding and memory benefits of schemas, as well as their costs. We return to the systematic problems that arise with expertise in the fifth section, Shortcomings of Expertise. Finally, expertise is domain specific. A person who is expert in money management may know to cut losses in investments but not in relationships (Tan & Yates, 1995). We consider the possibility of cross-domain decision expertise in the section titled General Decision Making.

The Role of the Environment in the Development of Expertise

One of the core findings in JDM research is that people are overconfident for many types of judgments and that there is a weak correspondence between statements of confidence about events and the actual occurrence of events (see Windschitl, Chapter 15 in this handbook). In predictions for which people are 90% confident, they are typically correct just 70% of the time. And a large number of studies have found that reputed experts are no better than novices at prediction and diagnosis. (Note that in these literatures, expertise is defined as a social construct – inferred from credentials or reputation and not from reproducible superior results. We return to problems that arise from reliance on social cues in the section Using Expertise). For example, early studies found that practicing clinical psychologists were no better than psychology students at making psychological diagnoses (Chapman & Chapman, 1969). Later reviews reached similar conclusions in regard to clinical diagnosis (Camerer & Johnson, 1991; Dawson et al., 1993) and in other domains, such as political predictions (Tetlock, 2005).

Despite a large number of discouraging examples, JDM has several celebrated examples of expert success, including weather forecasters, bridge players, and horse betters (see Bolger & Wright, 1994, for a review). For example, meteorologists are remarkably accurate (despite the skeptical stereotype about the profession). On occasions when meteorologists predicted that there was a 90% chance of rain, it actually rained 90% of the time (Murphy & Winkler, 1977). In terms of the overconfidence literature, weather forecasters were nearly perfectly calibrated. Their estimates corresponded to the true outcomes.[3] Experienced bridge players (Keren, 1987) and horse betters (Johnson & Bruce, 2001) have also been found to be near-perfectly calibrated in their likelihood judgments.

In contrast, clinical psychologists and political pundits are only marginally better than novices at predicting conditions and political events, respectively (Camerer & Johnson, 1991; Tetlock, 2005). Why are weather forecasters and bridge players more accurate than clinicians and political forecasters? There are two key factors that distinguish one set of activities from the other: the feedback environment in which people make predictions and learn outcomes, and the fundamental predictability of the task itself. For example, bridge players and weather forecasters receive rapid clear feedback on their estimates, whereas the outcomes that follow political forecasts are likely to be delayed in time and more ambiguous in their meaning. Weather forecasters and bridge players observe reliable cues that are readily available and consistent in their validity; political events, on the other hand, often have unique sets of conditions with no sample of analogous cases from which to generalize. And a clinical psychologist may never learn with certainty whether a diagnosed patient truly has a condition or not. The feedback environment and predictability of the task both play a critical role in the development of expertise, and we now consider each separately.

The feedback environment

One reason that weather forecasters can develop accurate judgments is that they are able to practice every day, and that practice is followed by clear, accurate, immediate feedback. Weather forecasters use available cues, such as changes in barometric pressure, wind patterns, and data on distant weather conditions, as well as historical patterns and models, to form a judgment about near-term weather, such as the likelihood of rain in a given locale in the next 24 hours. Within 24 hours of the prediction they find out whether it rained (using a consistent operationalization of the event). This is repeated day in and day out for years. Although unmotivated weather forecasters might evade or ignore feedback, professionalism and accountability will lead most weather forecasters to gauge their degree of success and to reflect on ways to improve their judgments.

Robin Hogarth (2001; see also Chapter 34 of this handbook) has proposed that some learning environments are kind, in that they provide clear, immediate feedback, which allows effective learning. The degree to which expertise development is possible depends largely on the kindness of the environment. Like weather forecasting, athletics take place in kind learning environments. Tennis players learn the appropriate conditions under which an unconventional shot, such as a lob or drop shot, will be effective through immediate feedback regarding each attempt's outcome. Experience in this case may be a slow teacher but evidence on a strategy's effectiveness is gathered quickly and clearly. However, not all environments are kind. Hogarth coined the term "wicked environment" to capture situations in which feedback is distorted, ambiguous, or delayed. Emergency-room physicians must make quick diagnoses and treatment decisions, yet the feedback they receive is incomplete – it captures an immediate outcome but not the full range of the long-term consequences. That delayed feedback, even if available, would be ambiguous because it would be affected by subsequent medical treatment and the behavior of the patient. Doctors also often observe a systematically biased subset of patients who return for future attention, which may skew their perception of the effectiveness of initial treatments.

There is a growing literature on learning environments that are wicked. Research shows that, in the absence of any feedback at all, people can hold highly inflated beliefs about themselves and their performance – but this tendency is reined in with the anticipation of immediate, accurate feedback (Armor & Sackett, 2006). Even when feedback is available, it may not be complete. One type of problem arises when choices early in a sequential feedback process preclude learning about other options (Denrell, 2005). For example, managers tend to learn the long-term performance of the people they hire but not the long-term performance of the job candidates they passed up (Einhorn & Hogarth, 1978). With asymmetric feedback on performance, it is difficult for a manager to evaluate whether the cues he or she is using to select new recruits are in fact predictive of success. Learning requires examining the covariation between cues (e.g., the job candidate has an Ivy League education or not) and subsequent job success and failure. A second type of incompleteness occurs when data are censored. For example, consider the information that retail managers receive when they are tasked with learning about the demand for their merchandise from past sales. If they do not sell all the merchandise they previously ordered, they get immediate and specific feedback regarding demand that month since they can observe the exact number sold and the number left over. However, if they *do* sell all the merchandise they previously ordered, they know that demand exceeded their inventory – but by how much? The clarity of feedback is asymmetric – the magnitude of the misestimate of demand is ambiguous when they sell out but clear when they do not. Over time, this pattern of asymmetric feedback leads to underestimation of true average demand even when judges are trained in the basic principles of supply chain management and are given incentives to be accurate (Feiler, Tong, & Larrick, 2013). Similar asymmetries in feedback can also cause individuals to form exaggerated risk perceptions, managers to underestimate the capabilities and trustworthiness of employees, and negotiators to develop inflated perceptions of their performance (see Feiler, Tong, & Larrick, 2013; Feiler, 2015; Larrick & Wu, 2007; Markle, 2011).

In sum, the kindest individual learning environments are those in which a person has a chance to make many decisions and receives rapid, clear, accurate feedback. There is a handy analogy for this kind of environment: it is a video-game. The beauty of a video game is that it is built on systematic regularities with salient cues and rapid, clear feedback, and it quickly provides a large sample of experience. In contrast, everyday life often gives people only a single try at a big decision and not hundreds. Moreover, video games encourage experimentation. An essential feature of learning is experimentation (or variation) in action taken coupled with feedback on outcomes

(Campbell, 1960; March, 1991). Video games lower the cost of failure. When you ill-advisedly leap for a distant platform and fall to your digital death, you immediately get a fresh start. In everyday life, however, people often experiment too little, both because they are overconfident in their understanding of their objectives and alternatives (Larrick, 2009) and because of the social costs of failing (Lee, Edmondson, Thomke, & Worline, 2004). As a consequence, individuals can become stuck in suboptimal ruts. Consider how easy it is for someone to develop the habit of driving the "usual way home" without contemplating other routes and how long they might take. Driving the presumed best-route precludes learning the driving time on the road not taken.

Wicked environments limit the possibility of developing expertise from personal experience. A second source of limitation is the fundamental predictability of the environment itself.

The predictability of the task

One of the classic traditions in JDM research is the Brunswikian approach to analyzing the performance of judges by constructing a model of the judge and a model of the task (Brunswik, 1952; Karelaia & Hogarth, 2008). A key term in the Brunswikian analysis is R_e, which represents the predictability of an outcome given the set of all available cues. Just as the reliability of a scale constrains its validity, the predictability of a task constrains the judgmental performance of judges over time. Consider an entirely unpredictable event: a fair coin flip. People may correctly guess a few coin flips, but this cannot be sustained over a large sample, and performance in one sample cannot predict performance in the next. Any superior performance in coin-flip anticipation is an illusion and not reproducible. A purely unpredictable task allows for no mastery or expertise because there is nothing to learn.

Thus, the second limit on the emergence of expertise is the degree to which a task is predictable. Weather patterns are predictable – a limited set of cues allow for accurate prediction. Medical problems are also predictable – a limited set of cues allow for accurate diagnosis. But many forms of human behavior are difficult to predict because they are multiply determined by many hidden factors. Shanteau (1992) argued that it is precisely these types of tasks on which there is little difference between novices and experts.

There is an interesting class of important tasks that occupy a gray middle ground. More accurately, they shift between black and white like an optical illusion. One version of the task is predictable and the other version is unpredictable, depending on how the task is defined. Consider the financial analyst who needs to value a company's stock. There are many cues to its value: current earnings, projected earnings, R&D investment, market competition, market growth, and so on. But a fairly simple rule captures the best prediction: The "right" stock price is the price that matches the company's price-to-earnings ratio (P/E) to the current P/E ratio in the market. There tends to emerge a wide consensus on the proper price for a company's stock. The difficult task is not *valuing* the stock – it is predicting *changes* in the stock price. Because markets are by and large efficient, changes in stock prices tend to follow a "random walk," that is, changes are unpredictable around the underlying value of the stock (Malkiel, 1973, 2003). Although superior "stock picking" performance can emerge in small samples, it is not reproducible over the long term. Thus, whether the task is predictable or not depends on whether one defines the role of the financial advisor as providing a fundamental value (predictable) or future changes in price (unpredictable). It is possible to be expert in valuation; it is extremely difficult to be expert in predicting price changes over large samples and over time.

Precisely the same analogy holds for sports markets: sports experts can accurately estimate the point spreads in football games, which correspond well to the actual margins of victory over a large sample. But sports experts cannot accurately predict which teams will beat the point spread. Similarly, the scouting staffs of National Football League (NFL) teams have some ability to assess the future talent of college players whom they are considering selecting in each year's draft. Players who are evaluated more highly, and drafted earlier, do in fact tend to perform better in their professional careers. But front offices do not differ reliably in their ability to select the best players. The reason: NFL teams invest enough in expertise that the ability to evaluate player talent is extremely high for all teams and varies little between teams. Although any NFL front office can outperform a group of novices in forecasting talent, they simply cannot consistently outperform each other. As a result, the pattern of team drafting success is exactly what would be expected in a random walk (Massey, 2011).

Thus, the question of whether there are financial experts or sports experts depends on the task you want the expert to perform. A financial expert can value a company and its stock and a sports pundit can assign a reasonable point spread to a game. Each of these numbers is predictable. But finding other valid cues that allow further prediction – of future stock price *changes* or which team will *beat* the spread – is effectively impossible. Competitive environments like financial investment, sports betting, and player drafting do depend on expertise to assess fundamental values but it is an arms race in which all parties acquire expertise to stay even with other parties. The remaining task – beating the market – is unpredictable and allows no expertise. True expertise, yielding reproducible, superior performance, is only possible in predictable tasks.

Individual versus collective learning

Generally, an analysis of learning environments focuses on whether the *individual* receives immediate, clear outcome feedback on his or her judgments. Fortunately, the development of expertise does not depend on rugged individuals extracting knowledge from the world all on their own. Even if individuals are confronted by a poor environment in which to learn, a collective of people can accumulate enough knowledge that individuals can "stand on the shoulders" of predecessors and learn rapidly (See top half of Figure 24.1). In these cases, individuals learn not through the brute force of personal trial and error but by being trained on accumulated wisdom and only then honing their newly acquired knowledge through deliberate practice. For instance, rather than forcing each medical student to learn individually how to treat ailments by simply encountering sick people, medicine harnesses collective learning by first teaching students how to treat ailments based on thousands of past studies. Only then do medical students begin interacting with patients to add depth and nuance to their understanding of medical practice. Thus, collective learning is an important complement to individual learning; a discussion of expertise that focuses only on individual processes paints an incomplete picture of human potential. Focusing on collective learning raises two important questions. What factors enable collectives to learn effectively (as a precursor to helping individuals learn more rapidly)? And how is this collective knowledge best conveyed to individuals?

Collectives learn because they expand the range of experience beyond what any one individual would encounter alone. For example, decentralized structures facilitate variation in the behavior of group members (Fang, Lee, & Schilling, 2010), which allows the collective to observe how different outcomes emerge from different strategies and then imitate the best ones. Cultural evolution is a form of collective learning by random variation, failure, selection, and retention across members of a society. For instance, our society has learned that some mushrooms are poisonous. This was presumably learned the hard way by brave (or foolhardy) ancestors who became ill or died when sampling the forest's hors d'oeuvres. Onlookers took note and added such trials to the collective bank of wisdom – "Don't eat the warty orange ones." The collective learned which mushrooms were safe and transferred that knowledge to subsequent generations of fungi consumers. Given the high cost of failure, this is not a friendly environment for "brute force" individual learning, making collective learning a more efficient approach.

Variation itself may be haphazard – as in cultural evolution – or can be approached more systematically, as exemplified in scientific exploration. The emergence of controlled experimentation and statistical analysis in the last 400 years has rapidly increased the rate at which collective wisdom has accrued. This is because it involves heightened intentionality. Rather than passively making associations, scientific exploration is a deliberative attempt to learn with the intention of capturing and sharing new knowledge.

A critical ingredient for successful collective learning is *continued* variation in order to keep up with the rapidly evolving world. A collective should always ensure that a few individuals seek to disconfirm conventions such that lessons are not accepted without sufficient testing or rendered suboptimal by changes in the environment. Unfortunately, history is chock-full of instances in which conventional wisdom failed. For centuries, doctors accepted bloodletting as a correct and effective treatment for illnesses, unaware that the convention was actually harmful for patients. Professional baseball teams overvalued players' speed relative to their hitting patience and power until the recent sabermetric revolution (Lewis, 2003; see Massey & Thaler, 2013 for a similar example in American football).[4] Similarly, the financial industry relied on a formula created by David Li called the Gaussian copula function to estimate the correlation in risk across different investments (Salmon, 2009). The function was elegant, implementable, and widely adopted. It was also overly simple. Therefore, the conventional approach to evaluating portfolio risk among financial experts ignored interdependencies that could arise in the real world, such as the "contagious" effects of mortgage failures as housing values crumbled and dragged down the value of whole segments of the housing market.

One final point on collective learning is that even once knowledge has been accumulated, successfully imparting it to individuals is both important and challenging. As shown

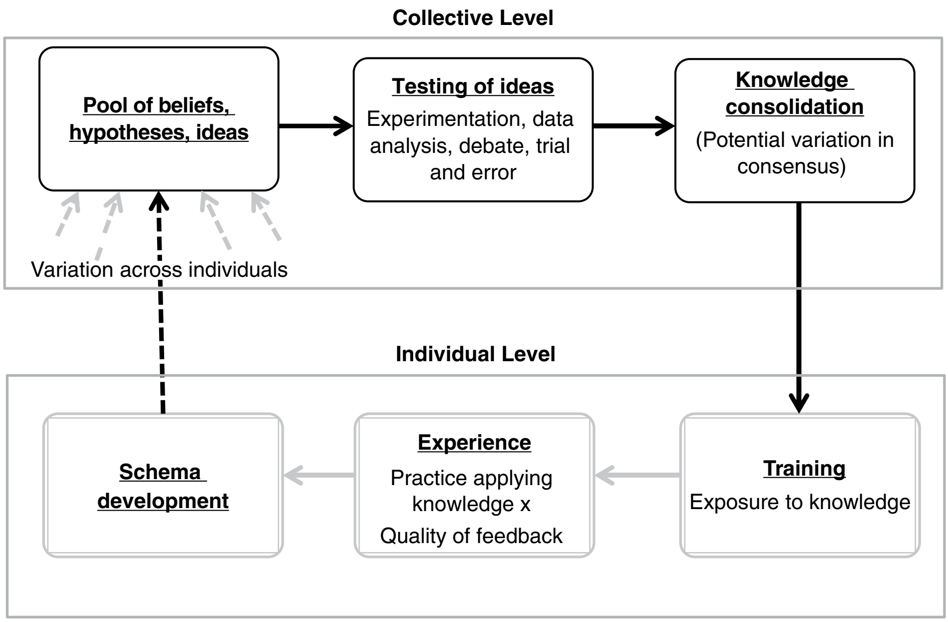

Figure 24.1 The generation of collective knowledge and transmission to individuals as a critical process in expertise development.

in Figure 24.1, knowledge accumulated at the collective level can be transmitted to the individual through training. But, the optimal process for individual learning goes beyond reading textbooks. As described in the section on the role of the environment, novices learn best if they can practice recognizing and using cues with clear, rapid feedback on the accuracy of their judgments and choices. Medical rounds involve taking textbook knowledge into the field to practice applying it; feedback comes in real time from a relative expert (the attending physician) and a less expert but knowledgeable collective (other students).[5] Situations that do not arise often in everyday practice can still be experienced and practiced vicariously by systematically studying rare cases in detail. Similarly, flight simulators help pilots-in-training practice applying their knowledge to a wide variety of circumstances without having to experience the costs of making a mistake.

Although optimal individual learning must go beyond reading textbooks, it is most effective if it builds on the kind of collective learning captured in textbooks. "Deliberate practice" (Ericsson, 2006) not only facilitates the development of expertise by ensuring a kind learning environment with rapid, accurate feedback but also ensures that people are attending to the right cues and relationships as identified through collective learning and as reflected in textbooks, classrooms, professional discussions, and so on. Unaided, people may focus on the wrong cues and will be susceptible to a number of memory biases as they try to learn from experience (Brehmer, 1980). By incorporating codified domain knowledge accumulated through collective learning (in formal fields such as medicine or chess), deliberate practice ensures that learners attend to the most valid cues and decision rules as they individually learn from experience.

In sum, collective learning is a vital complement to individual learning. Collective learning creates a body of veridical knowledge that can be transmitted to individuals (as shown in Figure 24.1), thereby giving each subsequent generation a head start and eliminating the need for each individual to "reinvent the wheel." A discussion of expertise that focuses only on individual processes of learning paints an incomplete picture of human potential. However, the same principles of individual expertise development still apply when trying to impart collective knowledge to individuals – the success of the transmission of collective wisdom depends on creating kind environments in which the individual can practice using the collective knowledge.

Is General Decision Making Expertise Possible?

While there is considerable evidence that systematic learning within a specific domain can yield superior, reproducible performance, there is little research on the existence of superior decision making performance across domains. A natural question for the field of JDM is whether studying or practicing it yields expertise that is general across domains? Can individuals *learn* to be more rational in their decision making? We consider two ways in which decision making expertise might be made generally applicable. The first is awareness: this approach assumes that people have the ability to learn to avoid the common decision traps studied in JDM research. The second is training: this approach assumes that a collection of superior decision making processes exist, and that they can be learned and applied across domains.

Expertise in avoiding decision traps

Considerable research has found that certain individual difference measures successfully predict performance on classic JDM tasks. Stanovich and West (1998) found that individuals who performed better on intelligence tests (Scholastic Aptitude Test, Raven Advanced Progressive Matrices, and Nelson–Denny Reading Test) were less prone to over confidence and the hindsight bias. Behavioral measures of cognitive reflection – the degree to which an individual's thinking is more deliberative and less impulsive – were predictive of more patient discount rates and less sensitivity to reference points in risky choice (Frederick, 2005). Parker and Fischhoff (2005) found that performance on decision tasks correlated positively with measures of introspective and analytical cognitive styles. In each of these cases, superior performance seems to be due to stable individual characteristics rather than developed skill.

Other evidence suggests that superior decision making could also be the product of development or learning. Bruine de Bruin, Parker, and Fischhoff (2012) found that although aging reduces cognitive fluidity, which hurts some decision making performance (e.g., susceptibility to framing and decision-rule use), it also provides experience and wisdom that improves performance in other decision making tasks (confidence calibration and resistance to sunk costs). Parker and Fischhoff (2005) found that one's social environment (e.g., peer environment and social support) positively correlates with performance on classic JDM tasks, even when controlling for cognitive ability. These results suggest that experience and social learning may provide valuable lessons that yield better performance on some decision tasks. Other research has looked at explicit training in basic rational principles of interest to JDM researchers. Nisbett and colleagues (1993) successfully used formal training sessions to increase rationality in decision making across domains. For instance, brief training sessions on the law of large numbers significantly improved the quality of statistical reasoning across a variety of everyday decisions (Fong, Krantz, & Nisbett, 1986). Some brief training on the principles of cost–benefit reasoning (e.g., ignoring sunk costs and accounting for opportunity costs) improved rationality in decision making even when the training was in a different context from decision making or when decision making occurred one month later (Larrick, Morgan, & Nisbett, 1990).

However, the development of such expertise from personal experience is likely to be difficult for two reasons. First, while decision making principles are general and abstract, real-world decisions are heavily laden with context. To infer patterns and develop expertise from experience one would need to draw connections between similar problems across disparate domains to extract the deeper structural commonalities, which makes learning of this type extremely difficult. Second, humans are renowned for their ability to interpret performance outcomes in a self-serving manner (Bradley, 1978). Learning is rendered impossible if success is considered to be the result of personal judgment while failure is considered to be the product of external factors. Without recognition of the commonalities across problems and accountability for good and bad performance alike, decision making expertise is unlikely to emerge from an individual's real-world experience.

Expertise in decision processes

Alternatively, an individual could be a general decision making expert through mastery of an effective decision process. Effective decision making typically consists of three phases (Heath & Heath, 2013; Keeney, 1996; Larrick, 2009; Payne, Bettman, & Schkade, 1999; Russo & Shoemaker, 2002): (a) structuring the problem (What are the objectives, alternatives, and possible outcomes?); (b) gathering information on outcomes (What are the benefits and costs of each possible outcome? How likely are those outcomes?); and (c) combining information to decide on an action (How should objectives be traded off? How much benefit is required to take a certain level of risk?). We suspect it is possible that people can develop a general expertise in the first and, to some extent, second stage. A decision making expert can aid the structuring of the problem across any domain by reducing the effects of framing and avoiding narrow-mindedness in the consideration of objectives and possible alternatives (Keeney, 1996). Expertise in decision making can also improve the gathering of information across many domains by facilitating the collection of representative samples, accounting for constraints in observed samples, and searching for evidence that can potentially disconfirm their beliefs.

Ultimately, however, domain-specific knowledge is needed for accessing key information and making informed trade-offs across attributes while weighing the risks. Without domain-specific knowledge, a decision maker does not know the relative importance of attributes or the validity of observed cues. Domain ignorance also leaves the decision maker blind to important interactions among factors that may be obvious to an individual experienced in that domain. We suspect that while an expert in decision processes can facilitate the structuring of the problem, consideration of all relevant costs and benefits, and the gathering of representative information, he or she will ultimately not have the domain-specific insights (e.g., the relevance and predictive validity of each piece of information) needed for effective decision making. This suggests that there is potential value for collaboration between a general decision making expert and a domain-specific expert. However, general decision making expertise has been little studied in the JDM literature and is a promising area to explore further.

Shortcomings of Expertise

The second section, Research on Expertise, reviewed many of the strengths of expertise that arise from possessing rich knowledge structures acquired through experience. However, knowledge and experience can also create shortcomings. We briefly review two here. The first is predictable memory shortcomings that arise from schematic processing. The second is an increased feeling of confidence that comes with expertise.

Schema-based shortcomings

Although schemas direct attention to relevant cues and actions, they also create systematic distortions in the processing of information. Schematic processing yields too little attention and memory for information unrelated to a schema (von Hippel, Jonides, Hilton, & Narayan, 1993). For example, a baseball fan may encode the ebb and flow of a baseball game but never give much attention to player uniforms or the location of concession stands. If some subsequent task would benefit from this information (e.g., creating a marketing campaign for the home team), the expert would be no better than a novice at offering advice on these dimensions. The flip side of paying too little attention to unrelated information is having false recall for schema-relevant information (Arkes & Harkness, 1980). Castel, McCabe, Roediger, and Heitman (2007) found that fans of professional football learned and recalled animal names more accurately if they were associated with professional teams (such as the falcons, colts, and bears); however, the fans also incorrectly recalled animal names that were not on the original list but that were also associated with professional teams (cardinals, panthers, and eagles).

Perhaps the main challenge of expertise is that it can lead to an entrenched way of thinking (Dane, 2010). High levels of expertise lead to "functional fixedness" so that all cases and decisions – routine or novel – are assimilated to prior ways of thinking. Entrenchment implies that experts may often perform well at generating incremental insights (that represent small changes on existing knowledge) but find generating radical insights more challenging (Dane, 2010). As long as the world is stable this can be quite effective; but, if new problems or opportunities come along, schemas can inhibit recognizing them. For example, Wood and Lynch (2002) showed that consumers with a high prior knowledge in a domain learned less about the features of a new allergy medicine than did those with a low degree of knowledge. More expert consumers acted as if they knew all there was to know and therefore did not attend closely to new information. However, if a "newness" cue was present, more knowledgeable consumers did attend closely to information and learned it more effectively than did less knowledgeable consumers.

Entrenchment also can give rise to difficulty on complex tasks that require multiple expert perspectives. Organizations exist precisely to tap the diverse range of expertise created by the division of labor. A diverse range of expertise has clear benefits: it increases the chances that someone in the organization can understand new, domain-specific technical knowledge and spot new, domain-specific market opportunities (Cohen & Levinthal, 1990). Many complex decisions benefit when multiple perspectives are integrated to capture a more complete set of relevant objectives. For example, cross-functional teams potentially benefit from the unique expertise of marketers, engineers, and financial analysts as they weigh trade-offs involving customer needs, manufacturing costs, and project financing. However, to the extent that each function has acquired its own schema for thinking about tasks, representational gaps between functions (Cronin & Weingart, 2007) can impede the ability of team members to agree on final decisions as each expert trumpets his or her own perspective.

In sum, schemas equip decision makers to attend to relevant information and to make rapid decisions. But they also come with shortcomings: they introduce their own distortions, including the neglect of schema-irrelevant information and an inability to shift one's thinking in new environments or when communicating with people who have a different expertise.

Overconfidence

Experts tend to be more accurate in their judgments than novices. They also tend to be more confident in their judgments (Ben-David, Graham, & Harvey, 2013; Tetlock, 2005). Ideally, experts would have enough insight to recognize their own limits such that they could be more aware of their uncertainty, in addition to being more accurate in their estimates. Many studies find, however, that better accuracy rarely comes with better calibration. For example, McKenzie, Liersch, and Yaniv (2008) found that, compared to undergraduate students, information technology (IT) experts gave more accurate estimates for IT-related questions (e.g., "As of January 2001, what percent of Americans used online banking services"), but their 90% confidence intervals were also narrower, expressing their greater confidence in their own ability to make predictions. As a result, the two groups were about equally overconfident as both IT experts and undergraduates gave 90% intervals that contained the true value less than 50% of the time. Thus, although experts made more accurate domain-specific estimates, they also were overly confident in their knowledge in the domain. These effects can be compounded when experts are unaware of the boundaries of their domain-specific expertise (Kahneman & Klein, 2009), in which case they suffer the greater overconfidence of expertise without the accuracy gain.

Using Expertise

As discussed in the third section, The Role of the Environment, expertise can only exist if an environment is, at least to some degree, predictable. As an environment increases in predictability, and in the presence of kind feedback, one can expect to gain from deferral to an expert. However, and somewhat counter intuitively, expertise can be further leveraged to improve upon experts themselves.

Extracting expertise to improve experts

Distilling expert knowledge into a decision support system can dramatically improve experts' consistency. Even very simple protocols, such as checklists, can improve performance by ensuring that important steps are not forgotten when working under pressure. For example, the state of Michigan was able to dramatically decrease the spread of infections in their intensive care units by instituting a sterility checklist for doctors to follow. The checklist was successful because it aided memory recall and made explicit the minimum necessary steps for effective treatment in such a challenging and complex environment (Gawande, 2009). By reducing variability and carelessness, even simple decision aids, which are informed by experts, can improve expert performance.

Similarly, with a large enough sample of a single expert's past judgments, one can build a model of how that expert used the cues in the environment (related to the aforementioned Brunswikian approach). This is known as a bootstrap model. Interestingly, research has found that such models can outperform the experts themselves on future predictions. Bootstrap models of experts are successful because they capture the wisdom of the expert while removing the random error of their intuitive judgment (Camerer, 1981; Hammond, Hursch, & Todd, 1964; Hoffman, Slovic, & Rorer, 1968; see also Dawes & Corrigan, 1974). Once again, such a methodology does not eliminate experts but, rather, taps into their expertise in a systematic way to reduce their future inconsistency.

Although JDM researchers appreciate the effectiveness of decision aids and expert models, practitioners and their clients are often skeptical (Arkes, Shaffer, & Medow, 2007; Kleinmuntz, 1990; See & Clemen, 2005). For example, doctors who rely on an electronic diagnostic support system are perceived to be less capable than doctors who make unaided diagnoses (Arkes, Shaffer, & Medow, 2007). We close by noting that there is a literature on the social and cognitive factors underlying technology acceptance (Venkatesh & Davis, 2000) but that the prejudice against unthinking formulas and algorithms is a barrier to harnessing these methods.

The difficulty of identifying experts

If one wants to *use* experts, then one must be able to *identify* experts. In practice, identification of real experts (as defined in the first section, Defining Expertise) is often remarkably difficult. In some instances, peers may readily agree on the identity of experts in their domain (Klein, 1998), which may be indicative of accumulated knowledge and possibly predictive of future performance. However, often such conceptions of expertise are socially constructed, for instance, through credentials or even personal expressions of confidence (Shanteau, 1992). These ways of conceptualizing expertise raise concerns about the relationship between the social cues to expertise (e.g., seniority and confidence) and the accuracy of future judgment. For example, research has found that people's stated confidence is only weakly correlated with performance (e.g., Burson, Larrick, & Klayman, 2006; Erev, Wallsten, & Budescu, 1994; Soll, 1996). The most confident expert will tend to perform better than others – but not nearly to the extent implied by their confidence (see Windschitl, Chapter 15 of this handbook). Given that many confident loudmouths perform poorly and many low-confidence individuals perform surprisingly well, should social constructions of expertise be relied on? We believe it is a risk.

Perhaps a simple answer is to rely on past or recent performance to identify experts. Unfortunately, people are usually faced with small samples of performance from a large pool of competitors. As a consequence, previous extremely high performance is often more likely to be the product of lucky risk taking than a signal of greater ability (Denrell & Liu, 2012). Denrell and Fang (2010) found that economists who make extreme predictions that turn out to be correct perform much worse in subsequent predictions. When samples of performance are small, then high performance is often a lucky match of erratic prediction with a random outcome and the performance advantage is not sustainable.

The case for the wisdom of select crowds

Many scholars have suggested that in the absence of sufficient data to identify real expertise, the optimal strategy may be to average all available judgments (Armstrong, 2001; Clemen, 1989; see Hastie & Kameda, 2005), thereby leveraging the "wisdom of the crowd" (Surowiecki, 2004). Averaging works because the collective is often centered near the truth, in which case high errors cancel low errors to yield a prediction near the truth. Even when the crowd is somewhat biased, averaging greatly reduces the variability in one's judgment, improving accuracy on average and greatly reducing the risk of an extreme error (Larrick & Soll, 2006).

But many people find averaging a crowd unattractive because it resigns itself to listening to the idiots in the crowd. A solution is to average a subset of the crowd. In recent work, Mannes, Larrick, and Soll (2014) have proposed an alternative strategy in which one forms a "select-crowd" of five judges based on whatever weak cues to expertise exist – such as one round of past performance. Through a battery of tests they find that this strategy is highly robust and performs particularly well as the range of expertise increases or as the crowd makes more independent errors. The select-crowd strategy performs well because it takes advantage of any signals to expertise (such as recent performance) but also reduces one's vulnerability to the random errors of any single individual. The ability to simultaneously leverage both the knowledge of experts and the reliability of a crowd is both intuitively and practically appealing.

Future Directions

We end the chapter by considering some general directions in which research on expertise in decision making might go next.

- **The interaction of collective knowledge and individual experience in expertise development.** When decision making researchers focus on expertise, they often focus on individuals needing to learn key relationships from experience, which is highly dependent on sample size, experimentation, and the quality of feedback. Fields of learning, however, such as medicine, replace the need for brute force individual learning with the ability to provide codified knowledge. But codified knowledge may be learned best by individuals when coupled with related experience. Many professional and disciplinary degree programs think that they are in the business of producing experts in those domains. An important question is whether the lessons of "textbook" learning only truly take hold when taught in conjunction with practice at application and whether they can be enhanced by systematic, well-designed experience.
- **The tragedy of the commonstance.** If expertise is built from the foundation of collective knowledge, then an interesting dilemma for collective learning is whether society properly incentivizes the challenging of conventional wisdom. We suggest that maybe, much like the *tragedy of the commons* – in which individuals deplete a common resource by pursuing their individual interests at the cost of the long-term best interests of the group – there is a *tragedy of the commonstance*: individuals may not challenge societal norms sufficiently because each individual prefers to conservatively employ the current "best practice," but as a whole the collective would be better off if we alternated testing, potentially disconfirming, and improving our conventional wisdom. Such a tragedy of the commonstance might be most costly for issues of public well-being, in which small changes for millions of people could yield big improvements in aggregate. We would also expect this problem to emerge for potential improvements that may not be patentable (and therefore not profitable). Hopefully, the recent establishment of governmental Behavioral Insights Groups in Great Britain and the United States will encourage the rigorous testing of conventional wisdom and best practices for the benefit of society.
- **The possibility of general expertise in decision making.** Can one become expert in decision making in general? Can the avoidance of biases or the use of better decision processes be learned and applied across domains? These questions are insufficiently studied in the JDM literature. Answering them would require the creation of tasks with which to learn good decision habits that can be presented across a wide range of contexts. The key would be not only to teach the possible pitfalls but also to increase cross-domain recognition of when to employ certain decision strategies. The possibility of creating the equivalent of medical rounds or flight simulators for practicing JDM insights is intriguing. Accumulated wisdom about generally effective decision practices would need to be translated into specific cases with which people can practice with accurate, immediate feedback in a kind learning environment. Unfortunately, many current courses

that teach decision making start and stop with demonstrations of biases. The ideas covered in the section on the role of the environment point in a new direction – they are a call to people who design curriculums to create more experiential ways of learning to make better decisions with an emphasis on problem recognition across situations. Recent research has demonstrated the benefits of training people using repetition coupled with accurate feedback to correct decision biases (Morewedge et al. (in press)) and to learn statistical relationships (Hogarth & Soyer, 2015).

The development and leveraging of expertise is fundamental to our progress as individuals, organizations, and a society. However, in a society in which business champions are crowned and felled as fast as markets can turn, individuals are also touted as experts one day only to disappoint the next. In this chapter we have outlined a framework for understanding what expertise is, when it is likely to emerge, and how it can be harnessed for sustainable superior performance.

Notes

1. Epstein (2013) provides an extensive, entertaining discussion of how some individual genetic differences – such as superior visual acuity for baseball players or an unusually long Achilles tendon for high jumpers – may allow the benefits of experience to accrue more rapidly for some athletes than for others. A similar difference may also arise for intellectual abilities (Campitelli & Gobet, 2011). Because the studies that have tried to measure the "number of hours" needed for expertise tend to start with an expert sample, they can underestimate the role that individual differences play in spawning expertise since they are sampling on the outcome. Less athletic or musical children may become discouraged from continuing with an activity, leaving those with more inherent talent to persist; in this selected group, hours of practice is the main factor that causes ultimate differences in ability.
2. The emphasis in this section on deliberate practices suggests that System 2 is heavily involved in the initial stages of expertise development. We note that learning often occurs automatically through associationistic processes – that is, through System 1 – in which case the quality of those associations is highly dependent on the nature of the feedback environment, as discussed in the next section, The Role of the Environment in the Development of Expertise.
3. It is worth noting that weather predictions cannot affect weather outcomes. Other predictions, such as stock predictions, can influence outcomes, increasing the correlation between predictions and outcomes but for potentially spurious (i.e., self-fulfilling) reasons.
4. Biases in evaluating prospective players in baseball easily persist because feedback is delayed and ambiguous. Prospects are usually evaluated as teenagers in high school and college. They then spend a number of years in the minor leagues as part of the development process. Predictions of ability are made many years before there is a clear evidence on actual ability, at which point many other factors have intervened (injuries, coaching, etc.) that make the link between prediction and outcome ambiguous. In the case of Major League Baseball, collective learning has been greatly facilitated by systematic analysis of data (Lewis, 2003) that goes beyond intuitive learning from feedback.

5. There are a number of additional techniques available to help experts make better decisions, such as using frequency formats instead of probabilities (Gigerenzer & Edwards, 2003), checklists (Gawande, 2009), and "nudges" such as defaults (Thaler & Sunstein, 2008). Also, see Soll, Milkman, and Payne, Chapter 33 of this handbook.

References

Arkes, H. R., & Harkness, A. R. (1980). Effect of making a diagnosis on subsequent recognition of symptoms. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 568.

Arkes, H. R., Shaffer, V. A., & Medow, M. A. (2007). Patients derogate physicians who use a computer-assisted diagnostic aid. *Medical Decision Making, 27*, 189–202.

Armor, D. A., & Sackett, A. M. (2006). Accuracy, error, and bias in predictions for real versus hypothetical events. *Journal of Personality and Social Psychology, 91*, 583–600.

Armstrong, J. S. (2001). *The principles of forecasting.* Norwell, MA: Kluwer.

Ben-David, I., Graham, J. R., & Harvey, C. R. (2013). Managerial miscalibration. *The Quarterly Journal of Economics, 128*, 1547–1584.

Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis. *Decision support systems, 11*, 1–24.

Bradley, G. W. (1978). Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology, 36*, 56–71.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 7717–7726.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica, 45*, 223–241.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2012). Explaining adult age differences in decision making competence. *Journal of Behavioral Decision Making, 25*, 352–360.

Brunswik, E. (1952). The conceptual framework of psychology. Chicago, IL: University of Chicago Press.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90*, 60–77.

Camerer, C. F. (1981). General conditions for the success of bootstrapping models. *Organizational Behavior and Human Performance, 27*, 411–422.

Camerer, C. F., & Johnson, E. J. (1991). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? *Toward a general theory of expertise: prospects and limits* (pp. 195–217). Cambridge, MA: Cambridge University Press.

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review, 67*, 380–400.

Campitelli, G., & Gobet, F. (2011). Deliberate practice is necessary but not sufficient. *Current Directions in Psychological Science, 20*, 280–285.

Castel, A. D., McCabe, D. P., Roediger, H. L. III, & Heitman, J. L. (2007). The dark side of expertise: Domain-specific memory errors. *Psychological Science, 18*, 3–5.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74*, 271–280.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 641–653.

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

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain related knowledge in relation to high or low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 257–273.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*, 559–583.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on innovation and learning. *Administrative Science Quarterly, 35*, 128–152.

Cronin, M. A., & Weingart, L. R. (2007). Representational gaps, information processing, and conflict in functionally diverse teams. *Academy of Management Review, 32*, 761–773.

Dane, E. (2010). Reconsidering the trade-off between expertise and flexibility: A cognitive entrenchment perspective. *Academy of Management Review, 35*, 579–603.

Dawes, R., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95–106.

Dawson, N. V., Connors, A. F., Speroff, T., Kemka, A., Shaw, P., & Arkes, H. R. (1993). Hemodynamic assessment in managing the critically III is physician confidence warranted? *Medical Decision Making, 13*, 258–266.

DeGroot, A. D. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.

Denrell, J. C. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112*, 951–978.

Denrell, J. C., & Fang, C. (2010). Predicting the next big thing: Success as a signal of poor judgment. *Management Science, 56*, 1653–1667.

Denrell, J. C., & Liu, C. (2012). Top performers are not the most impressive when extreme performance indicates unreliability. *Proceedings of the National Academy of Science, 109*, 9331–9336.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85*, 395–416.

Epstein, D. (2013). *The sports gene: Inside the science of extraordinary athletic performance*. Harmondsworth, UK: Penguin.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over-and under-confidence: The role of error in judgment processes. *Psychological Review, 101*, 519–528.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*, 570–581.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 39–68). New York, NY: Cambridge University Press.

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

Fang, C., Lee, J., & Schilling, M. A. (2010). Balancing exploration and exploitation through structural design: The isolation of subgroups and organizational design. *Organizational Science, 21*, 625–642.

Feiler, D. C. (2015). *The capability asymmetry: Why my boss underestimates me*. Working Paper, Dartmouth College.

Feiler, D. C., Tong, J. D., & Larrick, R. P. (2013). Biased judgment in censored environments, *Management Science, 59*, 573–591.

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, C. P. Feltovich, & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 41–67). Cambridge, UK: Cambridge University Press.

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.

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

Gawande, A. (2009). *The checklist manifesto: How to get things right*. New York, NY: Picador.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. Harmondsworth, UK: Penguin.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *BMJ: British Medical Journal, 327*, 741–744.

Gobet, F., & Simon, H. A. (1996). Recall of rapidly presented chess positions is a function of skill. *Psychonomic Bulletin Review, 3*, 159–163.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York, NY: Oxford University Press.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71*, 438–456.

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review, 112*, 494–508.

Heath, C., & Heath, D. (2013). *Decisive: How to make better choices in life and work*. New York, NY: Random House.

Hoffman, P. J., Slovic, P., & Rorer, L. G. (1968). An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. *Psychological Bulletin, 69*, 338–349.

Hogarth, R. M. (2001). *Educating intuition*. Chicago, IL: University of Chicago Press.

Hogarth, R. M., & Soyer, E. (2014). Using simulated experience to make sense of big data. *Sloan Management Review, 56*,49–54.

Johnson, J. E., & Bruce, A. C. (2001). Calibration of subjective probability judgments in a naturalistic setting. *Organizational Behavior and Human Decision Processes, 85*, 265–290.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American psychologist, 58*(9), 697.

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

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise. *American Psychologist, 64*, 515–526.

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*, 404–426.

Keeney, R. L. (1996). *Value-focused thinking: A path to creative decision making*. Cambridge, MA: Harvard University Press.

Keren, G. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior and Human Decision Processes, 39*, 98–114.

Keren, G., & Schul, Y. (2010). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science, 4*, 533–550.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, CT: Ablex.

Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin, 107*, 296.

Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2004). Using comprehensive feature lists to bias medical diagnosis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 563–572.

Larrick, R. P. (2009). Broaden the decision frame to make effective decisions. In E. A. Locke (Ed.), *Handbook of principles of organizational behavior* (pp. 461–480). Malden, MA: Blackwell.

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.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*, 111–127.

Larrick, R. P., Wu, G. (2007). Claiming a large slice of a small pie: Asymmetric disconfirmation in negotiation. *Journal of Personality and Social Psychology, 93*, 212–233.

Lee, F., Edmondson, A. C., Thomke, S., & Worline, M. (2004). The mixed effects of inconsistency on experimentation in organizations. *Organization Science, 15*, 310–326.

Lewis, M. (2003). *Moneyball: The art of winning an unfair game*. New York, NY: W.W. Norton.

Malkiel, B. G. (1973). *A random walk down Wall Street: The time-tested strategy for successful investing*. New York, NY: W.W. Norton.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives, 17*, 59–82.

Mannes, A. E., Larrick, R. P., & Soll, J. B. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology, 107*, 276–299.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*, 71–87.

Markle, A. B. (2011). Dysfunctional learning in decision processes: The case of employee reciprocity. *Strategic Management Journal, 32*, 1411–1425.

Massey, C. (2011). *The logic of chance*. Presented at the Institute for Operations Research and Management Science (INFORMS) conference, Charlotte, NC.

Massey, C., & Thaler, R. H. (2013). The Loser's Curse: Overconfidence versus market efficiency in the NFL draft. *Management Science*.

McKenzie, C. R., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes, 107*, 179–191.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C., Korris, J., & Kassam, K. S. (In press). Debiasing decisions: Improved decision making with a single training intervention. Policy Insights from the Behavioral and Brain Sciences. DOI: 10.1177/2372732215600886

Murphy, A. H., & Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Applied Statistics, 26*, 41–47.

Nisbett, R. E. (Ed.). (1993). *Rules for reasoning*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Norman, G. R., Trott, A. D., Brooks, L. R., & Smith, E. K. (1994). Cognitive differences in clinical reasoning related to postgraduate training. *Teaching and Learning in Medicine, 6*, 114–120.

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

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*, 96–116.

Payne, J. W., Bettman, J. R., & Schkade, D. A. (1999). Measuring constructed preferences: Towards a building code. *Journal of Risk and Uncertainty, 19*, 243–270.

Potter, S. (1947). *The theory and practice of gamesmanship, or the art of winning games without actually cheating*. London: Rupert Hart-Davis.

Russo, J. E., Schoemaker, P. J. (2002). *Winning decisions: Getting it right the first time*. New York, NY: Crown.

Salmon, F. (2009, February). Recipe for disaster: The formula that killed Wall Street. *Wired Magazine*, 23.

See, K. E., & Clemen, R. T. (2005). *Psychological and organizational factors influencing decision process innovation: The role of perceived threat to managerial power*. Working Paper, Duke University.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252–266.

Shanteau, J., & Stewart, T. R. (1992). Why study expert decision making? Some historical perspectives and comments, *Organizational Behavior and Human Decision Processes, 53*, 95–106.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22.

Slovic, P., Fleissner, D., & Bauman, W. S. (1972). Analyzing the use of information in investment decision making: A methodological proposal. *The Journal of Business, 45*, 283–301.

Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65*, 117–137.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161–188.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. London, UK: Little, Brown.

Tan, H. T., & Yates, J. F. (1995). Sunk cost effects: The influences of instruction and future return estimates. *Organizational Behavior and Human Decision Processes, 63*, 311–319.

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.

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

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science, 46*, 186–204.

von Hippel, W., Jonides, J., Hilton, J. L., Narayan, S. (1993). Inhibitory effect of schematic processing on perceptual coding. *Journal of Personality and Social Psychology, 64*, 921–935.

Wood, S. L., & Lynch, Jr., J. G. (2002). Prior knowledge and complacency in new product learning. *Journal of Consumer Research, 29*, 416–426.