

## 22

## EXPERT INTUITION: WHEN CAN WE TRUST IT?

Professional controversies bring out the worst in academics. Scientific journals occasionally publish exchanges, often beginning with someone's critique of another's research, followed by a reply and a rejoinder. I have always thought that these exchanges are a waste of time. Especially when the original critique is sharply worded, the reply and the rejoinder are often exercises in what I have called sarcasm for beginners and advanced sarcasm. The replies rarely concede anything to a biting critique, and it is almost unheard of for a rejoinder to admit that the original critique was misguided or erroneous in any way. On a few occasions I have responded to criticisms that I thought were grossly misleading, because a failure to respond can be interpreted as conceding error, but I have never found the hostile exchanges instructive. In search of another way to deal with disagreements, I have engaged in a few "adversarial collaborations," in which scholars who disagree on the science agree to write a jointly authored paper on their differences, and sometimes conduct research together. In especially tense situations, the research is moderated by an arbiter.

My most satisfying and productive adversarial collaboration was with Gary Klein, the intellectual leader of an association of scholars and practitioners who do not like the kind of work I do. They call themselves students of Naturalistic Decision Making, or NDM, and mostly work in organizations where they often study how experts work. The NDMers adamantly reject the focus on biases in the heuristics and biases approach. They criticize this model as overly concerned with failures and driven by artificial

experiments rather than by the study of real people doing things that matter. They are deeply skeptical about the value of using rigid algorithms to replace human judgment, and Paul Meehl is not among their heroes. Gary Klein has eloquently articulated this position over many years.

This is hardly the basis for a beautiful friendship, but there is more to the story. I had never believed that intuition is always misguided. I had also been a fan of Klein's studies of expertise in firefighters since I first saw a draft of a paper he wrote in the 1970s, and was impressed by his book *Sources of Power*, much of which analyzes how experienced professionals develop intuitive skills. I invited him to join in an effort to map the boundary that separates the marvels of intuition from its flaws. He was intrigued by the idea and we went ahead with the project—with no certainty that it would succeed. We set out to answer a specific question: When can you trust an experienced professional who claims to have an intuition? It was obvious that Klein would be more disposed to be trusting, and I would be more skeptical. But could we agree on principles for answering the general question?

Over seven or eight years we had many discussions, resolved many disagreements, almost blew up more than once, wrote many drafts, became friends, and eventually published a joint article with a title that tells the story: "Conditions for Intuitive Expertise: A Failure to Disagree." Indeed, we did not encounter real issues on which we disagreed—but we did not really agree.

## MARVELS AND FLAWS

Malcolm Gladwell's bestseller *Blink* appeared while Klein and I were working on the project, and it was reassuring to find ourselves in agreement about it. Gladwell's book opens with the memorable story of art experts faced with an object that is described as a magnificent example of a kouros, a sculpture of a striding boy. Several of the experts had strong visceral reactions: they felt in their gut that the statue was a fake but were not able to articulate what it was about it that made them uneasy. Everyone who read the book—millions did—remembers that story as a triumph of intuition. The experts agreed that they knew the sculpture was a fake without knowing how they knew—the very definition of intuition. The story appears to imply that a systematic search for the cue that guided the experts would have failed, but Klein and I both rejected that conclusion. From our point of view, such an inquiry was needed, and if it had been conducted properly (which Klein knows how to do), it would probably have succeeded.

Although many readers of the kouros example were surely drawn to an almost magical view of expert intuition, Gladwell himself does not hold that position. In a later chapter he describes a massive failure of intuition: Americans elected President Harding, whose only qualification for the position was that he perfectly looked the part. Square jawed and tall, he was the perfect image of a strong and decisive leader. People voted for someone who looked strong and decisive without any other reason to believe that he was. An intuitive prediction of how Harding would perform as president arose from substituting one question for another. A reader of this book should expect such an intuition to be held with confidence.

### INTUITION AS RECOGNITION

The early experiences that shaped Klein's views of intuition were starkly different from mine. My thinking was formed by observing the illusion of validity in myself and by reading Paul Meehl's demonstrations of the inferiority of clinical prediction. In contrast, Klein's views were shaped by his early studies of fireground commanders (the leaders of firefighting teams). He followed them as they fought fires and later interviewed the leader about his thoughts as he made decisions. As Klein described it in our joint article, he and his collaborators

investigated how the commanders could make good decisions without comparing options. The initial hypothesis was that commanders would restrict their analysis to only a pair of options, but that hypothesis proved to be incorrect. In fact, the commanders usually generated only a single option, and that was all they needed. They could draw on the repertoire of patterns that they had compiled during more than a decade of both real and virtual experience to identify a plausible option, which they considered first. They evaluated this option by mentally simulating it to see if it would work in the situation they were facing. . . . If the course of action they were considering seemed appropriate, they would implement it. If it had shortcomings, they would modify it. If they could not easily modify it, they would turn to the next most plausible option and run through the same procedure until an acceptable course of action was found.

Klein elaborated this description into a theory of decision making that he called the recognition-primed decision (RPD) model, which applies to firefighters but also describes expertise in other domains, including chess. The

process involves both System 1 and System 2. In the first phase, a tentative plan comes to mind by an automatic function of associative memory—System 1. The next phase is a deliberate process in which the plan is mentally simulated to check if it will work—an operation of System 2. The model of intuitive decision making as pattern recognition develops ideas presented some time ago by Herbert Simon, perhaps the only scholar who is recognized and admired as a hero and founding figure by all the competing clans and tribes in the study of decision making. I quoted Herbert Simon's definition of intuition in the introduction, but it will make more sense when I repeat it now: "The situation has provided a cue; this cue has given the expert access to information stored in memory, and the information provides the answer. Intuition is nothing more and nothing less than recognition."

This strong statement reduces the apparent magic of intuition to the everyday experience of memory. We marvel at the story of the firefighter who has a sudden urge to escape a burning house just before it collapses, because the firefighter knows the danger intuitively, "without knowing how he knows." However, we also do not know how we immediately know that a person we see as we enter a room is our friend Peter. The moral of Simon's remark is that the mystery of knowing without knowing is not a distinctive feature of intuition; it is the norm of mental life.

### ACQUIRING SKILL

How does the information that supports intuition get "stored in memory"? Certain types of intuitions are acquired very quickly. We have inherited from our ancestors a great facility to learn when to be afraid. Indeed, one experience is often sufficient to establish a long-term aversion and fear. Many of us have the visceral memory of a single dubious dish that still leaves us vaguely reluctant to return to a restaurant. All of us tense up when we approach a spot in which an unpleasant event occurred, even when there is no reason to expect it to happen again. For me, one such place is the ramp leading to the San Francisco airport, where years ago a driver in the throes of road rage followed me from the freeway, rolled down his window, and hurled obscenities at me. I never knew what caused his hatred, but I remember his voice whenever I reach that point on my way to the airport.

My memory of the airport incident is conscious and it fully explains the emotion that comes with it. On many occasions, however, you may feel uneasy in a particular place or when someone uses a particular turn of phrase without having a conscious memory of the triggering event. In hindsight,

you will label that unease an intuition if it is followed by a bad experience. This mode of emotional learning is closely related to what happened in Pavlov's famous conditioning experiments, in which the dogs learned to recognize the sound of the bell as a signal that food was coming. What Pavlov's dogs learned can be described as a learned hope. Learned fears are even more easily acquired.

Fear can also be learned—quite easily, in fact—by words rather than by experience. The fireman who had the “sixth sense” of danger had certainly had many occasions to discuss and think about types of fires he was not involved in, and to rehearse in his mind what the cues might be and how he should react. As I remember from experience, a young platoon commander with no experience of combat will tense up while leading troops through a narrowing ravine, because he was taught to identify the terrain as favoring an ambush. Little repetition is needed for learning.

Emotional learning may be quick, but what we consider as “expertise” usually takes a long time to develop. The acquisition of expertise in complex tasks such as high-level chess, professional basketball, or firefighting is intricate and slow because expertise in a domain is not a single skill but rather a large collection of miniskills. Chess is a good example. An expert player can understand a complex position at a glance, but it takes years to develop that level of ability. Studies of chess masters have shown that at least 10,000 hours of dedicated practice (about 6 years of playing chess 5 hours a day) are required to attain the highest levels of performance. During those hours of intense concentration, a serious chess player becomes familiar with thousands of configurations, each consisting of an arrangement of related pieces that can threaten or defend each other.

Learning high-level chess can be compared to learning to read. A first grader works hard at recognizing individual letters and assembling them into syllables and words, but a good adult reader perceives entire clauses. An expert reader has also acquired the ability to assemble familiar elements in a new pattern and can quickly “recognize” and correctly pronounce a word that she has never seen before. In chess, recurrent patterns of interacting pieces play the role of letters, and a chess position is a long word or a sentence.

A skilled reader who sees it for the first time will be able to read the opening stanza of Lewis Carroll’s “Jabberwocky” with perfect rhythm and intonation, as well as pleasure:

“Twas brillig, and the slithy toves  
Did gyre and gimble in the wabe:

All mimsy were the borogoves,  
And the mome raths outgrabe.

Acquiring expertise in chess is harder and slower than learning to read because there are many more letters in the “alphabet” of chess and because the “words” consist of many letters. After thousands of hours of practice, however, chess masters are able to read a chess situation at a glance. The few moves that come to their mind are almost always strong and sometimes creative. They can deal with a “word” they have never encountered, and they can find a new way to interpret a familiar one.

### THE ENVIRONMENT OF SKILL

Klein and I quickly found that we agreed both on the nature of intuitive skill and on how it is acquired. We still needed to agree on our key question: When can you trust a self-confident professional who claims to have an intuition?

We eventually concluded that our disagreement was due in part to the fact that we had different experts in mind. Klein had spent much time with fireground commanders, clinical nurses, and other professionals who have real expertise. I had spent more time thinking about clinicians, stock pickers, and political scientists trying to make unsupportable long-term forecasts. Not surprisingly, his default attitude was trust and respect; mine was skepticism. He was more willing to trust experts who claim an intuition because, as he told me, true experts know the limits of their knowledge. I argued that there are many pseudo-experts who have no idea that they do not know what they are doing (the illusion of validity), and that as a general proposition subjective confidence is commonly too high and often uninformative.

Earlier I traced people’s confidence in a belief to two related impressions: cognitive ease and coherence. We are confident when the story we tell ourselves comes easily to mind, with no contradiction and no competing scenario. But ease and coherence do not guarantee that a belief held with confidence is true. The associative machine is set to suppress doubt and to evoke ideas and information that are compatible with the currently dominant story. A mind that follows WYSIATI will achieve high confidence much too easily by ignoring what it does not know. It is therefore not surprising that many of us are prone to have high confidence in unfounded intuitions. Klein and I eventually agreed on an important principle: the

confidence that people have in their intuitions is not a reliable guide to their validity. In other words, do not trust anyone—including yourself—to tell you how much you should trust their judgment.

If subjective confidence is not to be trusted, how can we evaluate the probable validity of an intuitive judgment? When do judgments reflect true expertise? When do they display an illusion of validity? The answer comes from the two basic conditions for acquiring a skill:

- an environment that is sufficiently regular to be predictable
- an opportunity to learn these regularities through prolonged practice

When both these conditions are satisfied, intuitions are likely to be skilled. Chess is an extreme example of a regular environment, but bridge and poker also provide robust statistical regularities that can support skill. Physicians, nurses, athletes, and firefighters also face complex but fundamentally orderly situations. The accurate intuitions that Gary Klein has described are due to highly valid cues that the expert's System 1 has learned to use, even if System 2 has not learned to name them. In contrast, stock pickers and political scientists who make long-term forecasts operate in a zero-validity environment. Their failures reflect the basic unpredictability of the events that they try to forecast.

Some environments are worse than irregular. Robin Hogarth described "wicked" environments, in which professionals are likely to learn the wrong lessons from experience. He borrows from Lewis Thomas the example of a physician in the early twentieth century who often had intuitions about patients who were about to develop typhoid. Unfortunately, he tested his hunch by palpating the patient's tongue, without washing his hands between patients. When patient after patient became ill, the physician developed a sense of clinical infallibility. His predictions were accurate—but not because he was exercising professional intuition!

Meehl's clinicians were not inept and their failure was not due to lack of talent. They performed poorly because they were assigned tasks that did not have a simple solution. The clinicians' predicament was less extreme than the zero-validity environment of long-term political forecasting, but they operated in low-validity situations that did not allow high accuracy. We know this to be the case because the best statistical algorithms, although more accurate than human judges, were never very accurate. Indeed, the

studies by Meehl and his followers never produced a "smoking gun" demonstration, a case in which clinicians completely missed a highly valid cue that the algorithm detected. An extreme failure of this kind is unlikely because human learning is normally efficient. If a strong predictive cue exists, human observers will find it, given a decent opportunity to do so. Statistical algorithms greatly outdo humans in noisy environments for two reasons: they are more likely than human judges to detect weakly valid cues and much more likely to maintain a modest level of accuracy by using such cues consistently.

It is wrong to blame anyone for failing to forecast accurately in an unpredictable world. However, it seems fair to blame professionals for believing they can succeed in an impossible task. Claims for correct intuitions in an unpredictable situation are self-delusional at best, sometimes worse. In the absence of valid cues, intuitive "hits" are due either to luck or to lies. If you find this conclusion surprising, you still have a lingering belief that intuition is magic. Remember this rule: intuition cannot be trusted in the absence of stable regularities in the environment.

#### FEEDBACK AND PRACTICE

Some regularities in the environment are easier to discover and apply than others. Think of how you developed your style of using the brakes on your car. As you were mastering the skill of taking curves, you gradually learned when to let go of the accelerator and when and how hard to use the brakes. Curves differ, and the variability you experienced while learning ensures that you are now ready to brake at the right time and strength for any curve you encounter. The conditions for learning this skill are ideal, because you receive immediate and unambiguous feedback every time you go around a bend: the mild reward of a comfortable turn or the mild punishment of some difficulty in handling the car if you brake either too hard or not quite hard enough. The situations that face a harbor pilot maneuvering large ships are no less regular, but skill is much more difficult to acquire by sheer experience because of the long delay between actions and their noticeable outcomes. Whether professionals have a chance to develop intuitive expertise depends essentially on the quality and speed of feedback, as well as on sufficient opportunity to practice.

Expertise is not a single skill; it is a collection of skills, and the same professional may be highly expert in some of the tasks in her domain while remaining a novice in others. By the time chess players become experts,

they have "seen everything" (or almost everything), but chess is an exception in this regard. Surgeons can be much more proficient in some operations than in others. Furthermore, some aspects of any professional's tasks are much easier to learn than others. Psychotherapists have many opportunities to observe the immediate reactions of patients to what they say. The feedback enables them to develop the intuitive skill to find the words and the tone that will calm anger, forge confidence, or focus the patient's attention. On the other hand, therapists do not have a chance to identify which general treatment approach is most suitable for different patients. The feedback they receive from their patients' long-term outcomes is sparse, delayed, or (usually) nonexistent, and in any case too ambiguous to support learning from experience.

Among medical specialties, anesthesiologists benefit from good feedback, because the effects of their actions are likely to be quickly evident. In contrast, radiologists obtain little information about the accuracy of the diagnoses they make and about the pathologies they fail to detect. Anesthesiologists are therefore in a better position to develop useful intuitive skills. If an anesthesiologist says, "I have a feeling something is wrong," everyone in the operating room should be prepared for an emergency.

Here again, as in the case of subjective confidence, the experts may not know the limits of their expertise. An experienced psychotherapist knows that she is skilled in working out what is going on in her patient's mind and that she has good intuitions about what the patient will say next. It is tempting for her to conclude that she can also anticipate how well the patient will do next year, but this conclusion is not equally justified. Short-term anticipation and long-term forecasting are different tasks, and the therapist has had adequate opportunity to learn one but not the other. Similarly, a financial expert may have skills in many aspects of his trade but not in picking the stocks, and an expert in the Middle East knows many things but not the future. The clinical psychologist, the stock picker, and the pundit do have intuitive skills in some of their tasks, but they have not learned to identify the situations and the tasks in which intuition will betray them. The unrecognized limits of professional skill help explain why experts are often overconfident.

#### EVALUATING VALIDITY

At the end of our journey, Gary Klein and I agreed on a general answer<sup>13</sup> to our initial question: When can you trust an experienced professional? We

claims to have an intuition? Our conclusion was that for the most part it is possible to distinguish intuitions that are likely to be valid from those that are likely to be bogus. As in the judgment of whether a work of art is genuine or a fake, you will usually do better by focusing on its provenance than by looking at the piece itself. If the environment is sufficiently regular and if the judge has had a chance to learn its regularities, the associative machinery will recognize situations and generate quick and accurate predictions and decisions. You can trust someone's intuitions if these conditions are met.

Unfortunately, associative memory also generates subjectively compelling intuitions that are false. Anyone who has watched the chess progress of a talented youngster knows well that skill does not become perfect all at once, and that on the way to near perfection some mistakes are made with great confidence. When evaluating expert intuition you should always consider whether there was an adequate opportunity to learn the cues, even in a regular environment.

In a less regular, or low-validity, environment, the heuristics of judgment are invoked. System 1 is often able to produce quick answers to difficult questions by substitution, creating coherence where there is none. The question that is answered is not the one that was intended, but the answer is produced quickly and may be sufficiently plausible to pass the lax and lenient review of System 2. You may want to forecast the commercial future of a company, for example, and believe that this is what you are judging, while in fact your evaluation is dominated by your impressions of the energy and competence of its current executives. Because substitution occurs automatically, you often do not know the origin of a judgment that you (your System 2) endorse and adopt. If it is the only one that comes to mind, it may be subjectively undistinguishable from valid judgments that you make with expert confidence. This is why subjective confidence is not a good diagnostic of accuracy: judgments that answer the wrong question can also be made with high confidence.

You may be asking, Why didn't Gary Klein and I come up immediately with the idea of evaluating an expert's intuition by assessing the regularity of the environment and the expert's learning history—mostly setting aside the expert's confidence? And what did we think the answer could be? These are good questions because the contours of the solution were apparent from the beginning. We knew at the outset that fireground commanders and pediatric nurses would end up on one side of the boundary of valid intuitions and that the specialties studied by Meehl would be on the other, along with stock pickers and pundits.

It is difficult to reconstruct what it was that took us years, long hours of discussion, endless exchanges of drafts and hundreds of e-mails negotiating over words, and more than once almost giving up. But this is what always happens when a project ends reasonably well: once you understand the main conclusion, it seems it was always obvious.

As the title of our article suggests, Klein and I disagreed less than we had expected and accepted joint solutions of almost all the substantive issues that were raised. However, we also found that our early differences were more than an intellectual disagreement. We had different attitudes, emotions, and tastes, and those changed remarkably little over the years. This is most obvious in the facts that we find amusing and interesting. Klein still winces when the word *bias* is mentioned, and he still enjoys stories in which algorithms or formal procedures lead to obviously absurd decisions. I tend to view the occasional failures of algorithms as opportunities to improve them. On the other hand, I find more pleasure than Klein does in the comeuppance of arrogant experts who claim intuitive powers in zero-validity situations. In the long run, however, finding as much intellectual agreement as we did is surely more important than the persistent emotional differences that remained.

#### SPEAKING OF EXPERT INTUITION

"How much expertise does she have in this particular task? How much practice has she had?"

"Does he really believe that the environment of start-ups is sufficiently regular to justify an intuition that goes against the base rates?"

"She is very confident in her decision, but subjective confidence is a poor index of the accuracy of a judgment."

"Did he really have an opportunity to learn? How quick and how clear was the feedback he received on his judgments?"

## 23

### THE OUTSIDE VIEW

A few years after my collaboration with Amos began, I convinced some officials in the Israeli Ministry of Education of the need for a curriculum to teach judgment and decision making in high schools. The team that I assembled to design the curriculum and write a textbook for it included several experienced teachers, some of my psychology students, and Seymour Fox, then dean of the Hebrew University's School of Education, who was an expert in curriculum development.

After meeting every Friday afternoon for about a year, we had constructed a detailed outline of the syllabus, had written a couple of chapters, and had run a few sample lessons in the classroom. We all felt that we had made good progress. One day, as we were discussing procedures for estimating uncertain quantities, the idea of conducting an exercise occurred to me. I asked everyone to write down an estimate of how long it would take us to submit a finished draft of the textbook to the Ministry of Education. I was following a procedure that we already planned to incorporate into our curriculum: the proper way to elicit information from a group is not by starting with a public discussion but by confidentially collecting each person's judgment. This procedure makes better use of the knowledge available to members of the group than the common practice of open discussion. I collected the estimates and jotted the results on the blackboard. They were narrowly centered around two years; the low end was one and a half, the high end two and a half years.

Then I had another idea. I turned to Seymour, our curriculum expert,