

Can you have effective decision making without expertise? Review of G. Gigerenzer (2015).
Simply Rational: Decision making in the real world. Oxford University Press: Oxford, UK.
Review by Gary Klein, gklein@shadowboxtraining.com

It was a great pleasure to read Gigerenzer's *Simply Rational*, a collection of 12 of Gigerenzer's writings from 2007-2014. I have been following Gigerenzer's work for decades and have read several of these papers before, but it was useful to have them all together. There are so many insights and exciting ideas and practical recommendations.

The book is divided into 5 parts plus an introduction on how he got started teaching risk literacy to physicians and judges. But for practical purposes, there are three major sections: Risk communication (primarily in healthcare), Fast and Frugal Heuristics, and a critique of Behavioral Economics.

I will comment on each of these sections in order, but first a few general reactions:

- (a) The title appears to be a play on Dan Ariely's book "Predictably Irrational." Gigerenzer is providing a contrasting view: we are primarily rational, not irrational. Gigerenzer also describes some simple tactics we can use to improve decision making, such as methods for presenting data as frequencies instead of probabilities, and also some Fast and Frugal heuristics. These messages possess enormous significance because they contradict the current view that we are all crippled by biases.
- (b) In reviewing this book, I need to point out that I am not a decision researcher in the classical sense; I have conducted a very few controlled studies of decision making. And I have no background in economics. So it was a stretch for me to review Gigerenzer's efforts which blend decision research and economics. My background is in the field of expertise, and my comments below need to be understood from that perspective. The expertise lens helped me develop the Recognition-Primed Decision (RPD) model of how people actually make tough decisions under time pressure and uncertainty, as described in my book *Sources of Power: How people make decisions* (Klein, 1998, 20th anniversary issue published in 2018). I will have a lot more to say about expertise below.
- (c) A general comment about the writing — Gigerenzer is a very talented writer, but the format of collected papers raises an inevitable problem of redundancy because each paper must necessarily go over some of the ground covered by several of the others. This problem is a minor annoyance, and I don't know how it could have been avoided.

Now on to the review.

Risk Communication in Healthcare.

The first part of the book, about 40%, is about risk communication. This was my favorite set of articles. I could see a reason to present this material at the end, as a powerful close and a

demonstration of the practical implications of Gigerenzer's world view. However, I am glad he put it first because it is so important. The message is both frightening and inspiring. The frightening part is that well-respected decision makers, the physicians we rely on for life-and-death guidance, are typically clueless about how to interpret evidence about the risks and benefits of different types of treatment. The inspiring part is that this state of affairs can be remedied through reforms of the educational system and even through the formats we use to present evidence. Gigerenzer has been active in developing training programs in risk literacy and in conducting these programs; he has been active in devising more understandable formats for communicating benefits and risks, and has shown how effectively these work.

This body of work has profound implications and represents a major accomplishment. I am baffled that Gigerenzer's work on risk communication is not more widely known and put into practice. If a new vaccine for inoculating people against HIV was developed and approved, the headlines would quickly announce it and the demand would be immediate. Gigerenzer has been developing a cognitive vaccine for inoculating health care providers against risk confusions in interpreting the results of HIV tests, and for inoculating all of us against confusions and misinterpretations about risks and benefits of various treatments of all kinds. Yet this work has not received the prominence it deserves.

Consider judges studying DNA tests (p. 4). if you are accused of a murder and your DNA matches the traces on the victim, the prosecution might argue that the probability of a match by chance is 1 in 100,000. Sounds pretty damning. However, if you live in a city with 2,000,000 adult inhabitants, there will be around 20 adults whose DNA would match the traces on the victim. Not so damning after all. That's the kind of clarity that can emerge when we present data as frequencies rather than single-event probabilities. Gigerenzer goes into similar detail concerning mammograms, PSA tests for prostate cancer, and other issues.

Take breast cancer screening for women over the age of 50. For years the medical community has strongly recommended mammograms. And the data do show that the chance that a woman over 50 will die from breast cancer increases by 20% in the absence of screening. But let's look at the frequencies. Of 1,000 women with screening, 4 will die from breast cancer within 10 years. For 1,000 women who do not get regular screening, 5 will die of breast cancer within 10 years. Yes, that's a difference of 20% but it represents just one extra woman out of a thousand. And if you look at the frequency of dying of any type of cancer (because breast cancer can spread), then there is no difference between women who are screened for breast cancer and those who aren't. Then we consider the harms of screening: 100 out of 1,000 women who get screened will encounter false alarms, biopsies, and psychological distress. Five of the 1,000 women who get screened will have unnecessary treatments such as complete or partial breast removal even though their cancer is non-progressive. We have to look at the harms as well as the minimal benefits. That's why, according to Gigerenzer, the medical community is now backing away from its endorsement of breast cancer screening. However, it is not backing away as quickly as Gigerenzer leads us to believe. The American Cancer Society currently advises women to begin having mammograms at age 45, the U.S. Preventive Services

Task Force recommends age 50, and the American College of Obstetricians and Gynecologists says no later than 50.

Gigerenzer describes what it takes to achieve minimal statistical literacy in health, both for patients and for practitioners. It is not about mastering probability theory or Bayesian statistics. Rather, it is about becoming fluent in sorting through frequencies in order to think clearly about life-threatening conditions and potential treatments. It is about giving up delusions of certainty and accepting uncertainty. It is about insisting on reporting practices that make the benefits and the harms clear rather than disguising them. Gigerenzer has developed training programs, implemented them, and evaluated them to show that this type of training “sticks” — the training benefits persist even after the training itself ends.

Of course, the training should extend beyond healthcare professionals to include journalists and media specialists and legislators.

Even when the data show that people shouldn't be routinely getting mammograms and prostate cancer screening, segments of the medical community have been successful in promoting routine testing on the public relations front — witness the Senate voting 98-0 for a non-binding resolution in favor of mammography for women in their 40s, as happened about 20 years ago, despite the National Cancer Institute reporting insufficient evidence for this policy.

Despite my enthusiasm for this aspect of Gigerenzer's work, the science is not as clear-cut as he claims. Gilovich and Griffin (2002) have asserted that a number of studies on the benefits of frequency data have left a muddled picture — under some conditions frequency data lead to better judgments than probability data, but under other conditions, frequency data lead to worse decisions, and in other cases there is no real difference. Therefore, Gigerenzer needs to explain the boundary conditions for when frequency data are most appropriate. None of the Gigerenzer papers in this section of *Simply Rational* mention the Gilovich and Griffin claim, even though all of them were published after the Gilovich and Griffin chapter appeared. Gigerenzer himself has a chapter in *Heuristics and Biases* (Gilovich et al., 2002), the book that the Gilovich and Griffin critique appeared (it was the introductory chapter in that book). It is entirely possible that Gigerenzer has addressed this issue elsewhere — if so I would have appreciated having that paper included in this collection. For now, I assume Gigerenzer has learned the conditions under which frequency data make a difference.

In places, Gigerenzer goes a bit further than I am comfortable with. For example on p. 55 he comes out strongly in favor of evidence-based medicine (EBM), ignoring the serious cognitive challenges of implementing EBM (e.g., D. Klein, Woods, G. Klein, Perry, 2016; 2018). And while this seems like a minor quibble, it is symptomatic of a deeper concern I have with Gigerenzer's work — the way it ignores domain expertise. This concern comes up more forcefully with regard to the next section of *Simply Rational* on smart heuristics.

Smart Heuristics

One of the major takeaways from this section is that we should not automatically accept the analytical standards for making judgments and decisions. Gigerenzer explains how the requirements, such as having large samples and clear data, are typically not met in natural settings because these are uncertain and ambiguous.

Gigerenzer also questions the claim that heuristics are invariably a source of errors. His research has shown that simple heuristics can be as accurate as formal analyses, and even more accurate in certain conditions. For example, the gaze heuristic allows us to perform difficult tracking tasks without having to do any analyses at all. When pursuing an object (e.g., a bird attacking a prey or a baseball outfielder trying to catch a fly ball), one tries to maintain a constant optical angle between the target and oneself, rather than estimating where the target is likely to wind up and then performing the mental calculations to compute a trajectory to arrive at that point (Hamlin, 2017).

I particularly like the methodological principle Gigerenzer offers in that in conducting research we should not base conclusions on averages alone but should examine individual data. I think there are lot of important discoveries to be made in pondering the responses of individual participants in a study. This principle was drummed into my head back in graduate school in a course on Sensation/Perception, discussing the rod-cone break in visual sensitivity. When researchers measured visual sensitivity while dark-adapting their participants and averaged the data, they found a smooth curve. We become more sensitive to stimuli the longer we are dark adapted. However, none of the individual participants actually showed a smooth curve. Each one showed a different pattern, an increase in sensitivity for the first few minutes, then an asymptote, and then after another few minutes another increase in sensitivity followed by a second asymptote. This pattern revealed the rod system kicking in after several minutes of dark adaptation. Yet every research participant had his/her own pattern, with the rod/cone break showing up faster in some and slower in others, so averaging the data smoothed out the inflection point and hid the actual phenomenon.

Similarly, I suspect we lose a lot of data by ignoring outliers and individual tendencies in research on judgments and decisions. When I have gone back to look at the Kahneman and Tversky studies demonstrating biases, the data suggest that 10-20% of the participants do not show the errors, which should caution even the advocates of judgment biases against claiming that the biases are universal.

The section on Smart Heuristics raises some issues on which my views diverge from those of Gigerenzer. The most fundamental difference is how we view expertise (Klein, 2015). For Gigerenzer, expertise depends on calibrating with the cue structure in the environment, whereas I identify experts as people with rich repertoires of patterns. Gigerenzer seeks to study the degree of calibration whereas I study experts by using cognitive task analysis methods to examine their mental models. Gigerenzer is rightfully concerned with way traditional researchers disregard heuristics whereas I am also concerned with the way our society denigrates expertise itself.

To that end, my colleagues and I have written about the war on expertise that is being waged on several fronts: the decision research community asserting that algorithms can out-perform experts, the evidence-based practice community arguing that practitioners should rely on the results of controlled experiments rather than on their experience, the computer science community claiming that Artificial Intelligence systems can beat experts, the sociology community asserting that expertise is a social construct rather than an individual one, and the heuristics and biases community demonstrating that even experts are prone to biased judgments (Klein et al., 2018). All of these assertions and claims are misleading and inaccurate and in my opinion we need to be very careful to acknowledge the importance of expertise, especially when performing tasks in complex and ambiguous situations.

My problem is that I don't see much scope for expertise within Gigerenzer's Fast and Frugal Heuristics framework. In contrast, the Naturalistic Decision Making (NDM) community in which I work is fascinated with experts and sees the study of expertise as one of the central features of NDM projects.

How might one increase expertise: Gigerenzer might try different methods to use the heuristics in the adaptive toolbox he is developing. However, I prefer to use methods such as ShadowBox (Klein & Borders, 2016) to build richer mental models. ShadowBox is a scenario-based approach that lets trainees see the world, or at least the scenario, through the eyes of experts. We have shown that even in a half-day of training, people achieve a 20-25% stronger match to the experts.

The difficulty here is not just that Gigerenzer neglects expertise. One of Gigerenzer's favorite heuristics, the Recognition Heuristic, works better when we know less! Therefore, expertise gets in the way of using this heuristic.

And that brings me to a second concern, the restrictive conditions for using some of the heuristics Gigerenzer describes. The Recognition Heuristic takes advantage of my lack of knowledge. For example, American students are more accurate at gauging the size of German cities than German students because the Americans can rely on familiarity — the more familiar a German city, the more likely it is to be larger. The German students don't have this cue because they had heard of all the cities. Gigerenzer's team has demonstrated the advantages of the Recognition Heuristic for almost two decades.

However, I am wondering how I would use the Recognition heuristic. If I have zero expertise — let's say I lived all my life in a remote Amazon village — I couldn't use the Recognition heuristic because I wouldn't have any familiarity. If I grew up in Germany, I would have too much expertise. How do I judge for myself if I have just the right amount of expertise to enable the Recognition Heuristic to work? Similarly, the Take the Best heuristic requires that I can sort cues by their validity, which sounds straightforward until you try it.

Therefore, I am left wondering about how applicable and how practical the heuristics in the Adaptive Toolbox really are.

A third type of misgiving I have is that Gigerenzer seems quick to dismiss his intellectual adversaries and their accomplishments, rather than trying to see what he can learn from them. This stance may be necessary when articulating a bold vision, as Gigerenzer is doing, but it does not always result in productive inquiry.

Consider Gigerenzer's attacks on the Kahneman and Tversky work on the availability and representativeness heuristics because these do not adhere to the criteria Gigerenzer has advocated. However, the Kahneman and Tversky research was conducted two decades before Gigerenzer articulated his criteria, so it seems odd to apply them retroactively in this way. I would have preferred to see Gigerenzer acknowledging the great value that Kahneman and Tversky provided by the way they took Herbert Simon's general notion of heuristics and identified several specific examples, such as availability and representativeness. I wonder how far Gigerenzer would have gotten with his own investigation of heuristics had it not been for the groundbreaking work of Kahneman and Tversky.

Consider also Gigerenzer's advocacy of a fluency heuristic, which seems somewhat similar to the availability heuristic described by Kahneman and Tversky. Doubtless there are some points of distinction but I think there are also major points of convergence.

And while I admire Gigerenzer's demonstrations of the power of using frequency data, he might have pointed out that Kahneman and Tversky (1982; Tversky and Kahneman, 1983) had themselves explained that under certain conditions people could be more accurate using frequency data than probabilities.

All of this leaves me wondering if Gigerenzer may be in some ways trapped by his arguments with the Heuristics and Biases community and with the larger Judgment and Decision Making community, performing studies that are rebuttals to their ideas rather than moving on in new directions. I would like to see how the notions of expertise I discussed earlier might get incorporated into the Adaptive Toolbox. I would like to see Gigerenzer explore ways to help people use subtle cues and tacit knowledge in addition to his concern for explicit cues.

My fourth and final concern is that some of Gigerenzer's ideas have precedents that he might not be aware of. For example, on p. 132 he describes a method of tallying that simply counts the number of cues favoring one alternative to the others. Benjamin Franklin identified this method several centuries ago. Next, on p. 143 he discusses the potential value of forgetting — I think he might enjoy reading Schacter's book *The Seven Sins of Memory* (2001), which describes the ways that memory failures actually have great value. On a personal level, I noted that on page 122 Gigerenzer cited the work of Johnson & Raab (2003) showing the use of the Take-the-First heuristic; I would suggest he look at Klein et al. (1995) demonstrating that phenomenon almost a decade earlier. And finally, on p. 136 Gigerenzer discusses the value of a default rule for making decisions, but curiously he does not acknowledge *Nudge* (2008), in

which Thaler and Sunstein describe their applications of a default rule. Gigerenzer discusses Thaler and Sunstein in the section on Behavioral Economics, but only to criticize them rather than to acknowledge their contribution. So let's turn next to that section of *Simply Rational*.

There is an intervening section with chapters on the hot hand effect and on gender stereotypes regarding intuitions, but these chapters don't have the heft of the others and seem to be serving the function of filling out the book more than anything else

Behavioral Economics

As might be expected, Gigerenzer is not enthusiastic about behavioral economics and raises some important concerns. He is troubled by the way the researchers emphasize fitting the data rather than on making predictions. Fitting the data is an interesting intellectual exercise but the real impact comes when scientists try to make more accurate predictions or at least try to offer genuine explanations for the data.

Gigerenzer points out that behavioral economists have thus far provided no evidence that people who deviate from axiomatic rationality and neoclassical norms suffer for their cognitive transgressions. They don't show lower earnings, lower happiness, impaired health, or shorter lives. He urges the research community to study successful decision makers to try to understand what makes them so good. And he reiterates his misgivings about the belief that all observed actions result from a constrained optimization process. He provides an amusing example of how a commensurability analysis can go awry — consider a person trying to buy a house using bundles of features to be traded off (e.g., square footage, price, number of bathrooms, and so forth). Such a strategy can lead to a situation in which a sufficiently high number of bathrooms could compensate for a minuscule house size.

Gigerenzer also makes the important observation that people who are perfectly consistent will score well on measures of rationality — in fact, neoclassical economists define “perfect internal consistency as *the standard of rationality*” (p. 246). Gigerenzer notes that this position is odd because someone may be mistaken about everything of consequence and may be completely inaccurate in judgments, and still get high marks because of consistency. It may be that without reliability, one cannot have valid judgments. Nevertheless, reliability and internal consistency do not imply valid and accurate judgments. Someone who is reliably wrong is still wrong. Internal consistency is highly valued by the Judgment and Decision Making community for many reasons, not least of which is that it is so easy to measure.

Although I am very sympathetic to Gigerenzer's arguments in this section, I kept waiting for something to happen that never did: Gigerenzer never mentions the contributions for which behavioral economics is most well known — the methods for influencing judgments and decisions. Here is the only comment Gigerenzer makes to that end, in a 26-page long section about behavioral economics: “Behavioral economists who decades ago pitched the behavioral approach to the neoclassical mainstream as a purely descriptive enterprise (e.g., Tversky & Kahneman, 1986; Thaler, 1991, Frank, 1991 — and nearly everyone else published in top-

ranked economics journals), now advocate using behavioral concepts for prescriptive policy purposes (Thaler & Sunstein, 2008; Frank, 2008; Amir et al., 2005)." And that's it.

I don't understand how Gigerenzer can avoid acknowledging the tremendous practical impact that behavioral economics has had. Its practitioners are eager to formulate training and public policy guidelines for influencing the populace, just as Gigerenzer has attempted to do regarding risk communication and described in the first section of the book. To be sure, many of the tactics adopted such as the use of social proof that are used by nudge units in various countries come out of applied social psychology rather than the heuristics and biases research (e.g., Cialdini, 1993; Wilson, 2011). Nevertheless, the practical applications have gained a world-wide following that is quite impressive and some of the primary levers for gaining voluntary compliance do reflect findings from heuristics and biases researchers, e.g., loss aversion, framing, and anchoring and adjustment.

I wish Gigerenzer could gain some satisfaction from seeing how these "nudge" tactics have spread, and I hope that his valuable work on risk communication will gain the same kind of popular acceptance.

References

- Cialdini, R.B. (1993). *Influence: Science and Practice*. Boston, MA: Allyn & Bacon.
- Gilovich, T., & Griffin, D. (2002). Introduction — Heuristics and biases: Then and now. In T. Gilovich, D. Griffin & D. Kahneman (Eds.). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.
- Hamlin, R.P. (2017). "The Gaze Heuristic:" Biography of an adaptively rational decision process. *Topics in Cognitive Science*, 9, 264-288.
- Kahneman, D. & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, 11, 123-141.
- Klein, D. E., Woods, D. D., Klein, G., & Perry, S. J. (2016). Can we trust best practices? Six cognitive challenges of evidence-based approaches. *Journal of Cognitive Engineering and Decision Making*, 10(3), 244-254.
- Klein, D., Woods, D., Klein, G., & Perry, S. (2018). EBM: Rationalist fever dreams. *Journal of Cognitive Engineering and Decision Making*, 12(3), 227-230.
- Klein, G. (2015). A naturalistic decision making perspective on studying intuitive decision making. *Journal of Applied Research in Memory and Cognition*, 4(3), 164-168.

Klein, G., & Borders, J. (2016). The ShadowBox approach to cognitive skills training. *Journal of Cognitive Engineering and Decision Making*, 10(3), 268-280

Klein, G., Shneiderman, B., Hoffman, R.R., & Wears, R.L. (2018). The “war” on expertise: Five communities that seek to discredit experts. In P. Ward, J.M. Schraagen, J. Gore, and E.M. Roth (Eds.) *The Oxford Handbook on Expertise*. Oxford: Oxford University Press.

Schachter, D. (2001). The seven sins of memory: How the mind forgets and remembers. New York: Houghton Mifflin Harcourt.

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

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293-315.

Wilson, T.D. (2011). Redirect: The surprising new science of psychological change. Columbus, GA: Little, Brown & Company.