Disclaimer

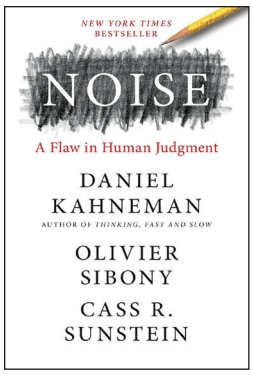
Klaas Sijtsma is my collaborator and colleague. Had I disliked the book, I would not have reviewed it.

About the Author

Eric-Jan Wagenmakers is a professor of Bayesian methodology at the University of Amsterdam and an enthusiastic supporter of transparent statistical reporting. His lab guides the development of the open-source statistical software program JASP (*jasp-stats.org*).

Noise: A Flaw in Human Judgment

Daniel Kahneman, Olivier Sibony, and Cass R. Sunstein



Paperback: 460 pages

Publisher: Little, Brown and Company, 2021

ISBN-10: 0316451398

Reviewed by Gaurav Sood and Andrew Gelman

It's natural to think of judgment in terms of mathematical functions in which the same inputs map to the any human decision-making systems. Take insurance underwriting, for instance. Given the same data (realistic but made-up information about cases), the median percentage difference between quotes for any pair of underwriters is a stunningly large 55% (so for half of the cases, it is worse than 55%), a difference about five times as large as expected by the executives asked about this scenario in a survey.

Several points flow from this result. If you are a customer, your optimal strategy is to get multiple quotes. What explains the ignorance about the disagreement? There could be a few reasons. First, when people come across a quote from another underwriter, they may "anchor" their estimate on the number they see, reducing the gap between the number and the counterfactual. Second, colleagues plausibly read to agree, asking, "Could this make sense?" rather than, "Does this make sense?"

The consequences of the mismatch between naïve deterministic models and real-world individual and social decision processes are the topic of the recent book *Noise: A Flaw in Human Judgment*, by psychologist Daniel Kahneman, business strategist Olivier Sibony, and law professor Cass R. Sunstein. *Noise* is an appropriate follow-up to earlier books by Kahneman on heuristics and biases in human judgment and by Sunstein on legal and social institutions. Kahneman et al. discuss a study from asylum reviews, writing, "A study of cases that were randomly allotted to different judges found that one judge admitted 5% of applicants, while another admitted 88%."

Variability can stem from two things. Maybe the data doesn't allow for a unique judgment (irreducible error). But even here, the final judgment should reflect the uncertainty in the data or that at least one person is disagreeing with the consensus, which can arise from variation in skill (e.g., how to assess visa applications), variation in effort (e.g., some judges put more effort than others), agency and preferences (e.g., I am a conservative judge, and I can deny an asylum application because I have the power to do so), or biases induced by cognitive errors or the use of irrelevant information (e.g., weather, hypoglycemia, etc.).

A lack of variability doesn't mean we have the right answer, but the existence of variability puts the lie to implicit deterministic models of the social world. As quantitative social scientists, we appreciate the message of Kahneman et al. in this book. The concept of noise—in statistical terms, unexplained variation—is central to our work and our understanding of empirical research, but we do not always think about how this interacts with what might be called the folk psychology of human decision-making. Their point is not just that noise exists but that we need to confront our deterministic intuitions.

Mismatch between Claims and Research Method

Ironically, *Noise* makes some errors it warns about. Here are two patterns we noticed, that follow a general practice of not questioning results that are congenial to the story:

1. Extremely small n studies cited without qualification. For example, the authors note, "when the same software developers were asked on two separate days to estimate the completion time for the same task, the hours they projected differed by 71%, on average." The cited study was based on only seven developers. In a discussion

of hypoglycemia and judgment, Kahneman et al. cite a paper on "what the judge ate for breakfast" that used data from only eight judges.

2. Surprising but likely unreplicable results. For example, the authors write, "When calories are on the left, consumers receive that information first and evidently think 'a lot of calories!' or 'not so many calories!' before they see the item. Their initial positive or negative reaction greatly affects their choices. By contrast, when people see the food item first, they apparently think 'delicious!' or 'not so great' before they see the calorie label. Here again, their initial reaction greatly affects their choices. This hypothesis is supported by the authors' finding that for Hebrew speakers, who read right to left, the calorie label has a significantly larger impact." As explained by Francis and Thunell in "Excess Success in 'Don't Count Calorie Labeling Out: Calorie Counts on the Left Side of Menu Items Lead to Lower Calorie Food Choices'", Dallas et al.'s research in "Don't Count Calorie Labeling Out" has several problems that make us doubt its claims would replicate or apply in the real world.

We see a fundamental incoherence between *Noise*'s key substantive point—that, contrary to intuition, individual and group decisions are noisy, or fundamentally unpredictable—and its research method, which is to interpret studies as implying something close to universal truths. To put it another way, if software developers' judgments are noisy, and if judges' decisions are easily swayed by irrelevant factors, then how could we expect to extract general insights from a small study of seven or eight people at one place and time?

A useful implication we have drawn from the general thesis of *Noise* is that our awareness of the unpredictability of individual and social decisions should make us skeptical of naïve expectations of predictability, even in a study of insurance underwriting where there would seem to be clear incentives for economic efficiency, and also skeptical of simple solutions backed by unreplicated research. Noise can be measured and possibly reduced in level, but it cannot be hacked easily.

The Conception of the Book

The study of noise and uncontrolled variation has a long history in statistics and economics, but the authors of this new book are neither statisticians nor economists. They came at the topic as outsiders. In "This Hidden Flaw Can Lead to Grave Errors in Economic Predictions" by Jane Wollman Rusoff, Sunstein is quoted saying, "Unlike bias, noise isn't intuitive, which is why we think we've discovered a new continent."

From the standpoint of statistics, economics, or quality engineering, noise is not a new idea at all—so Sunstein's previous unfamiliarity with the topic suggests statisticians and economists have failed to fully communicate this idea to the public.

Sunstein also says, "One [of the things] I learned in this [book] collaboration is not to think in terms of [for instance], 'Will this stock go up'? 'Is this the right investment strategy?' but instead to think: 'What's the probability that this stock will go up?' 'What's the probability that this is a good investment strategy?' So rather than asking, 'Is it good to invest in international stocks [versus] domestic stocks?', it's better to ask, 'What probability do you assign to the proposition that international stocks will outperform domestic stocks in 2022?'"

We agree it is a good idea to think probabilistically, but this also seems like common sense. Don't financial advisers tell you all the time that we can't know the future? We can only guess and at best assign probabilities. We are reminded of that scene by academic satirist David Lodge where a group of English professors is sitting in a circle playing a game where they take turns listing famous books that, embarrassingly, they've never read, and one of them lists Hamlet. A bit too embarrassing, it turns out! Similarly, it's kind of admirable how open Sunstein is about his former cluelessness, but it makes you wonder whether he was the most qualified person to write a book about a topic that lots of people know about but he'd never thought about until five years ago.

Also, a minor point is that we don't think it's quite right to ask questions like "What's the probability that this stock will go up?" Sure, you can ask the question to check that your investment adviser is on the ball, but it doesn't make sense to think of the stock price going up or down as a binary outcome. The investment adviser should be thinking of things like expectation and tail risk. It's not a big deal but perhaps revealing of the authors' continuing discomfort with the concept of noise.

Noise discusses important issues connecting cognitive heuristics and biases to our understanding of decision-making. It is an interesting book that relates to, but goes beyond, its authors' earlier influential work on cognitive psychology and social processes. The book has a weakness in that its interpretation of research follows a fallacy that it warns us about in other contexts, which is to interpret unreplicable—noisy—research claims as implying an unrealistic predictability about the social world. This mistake may stem from the authors' unfamiliarity with the existing understanding of noise

in statistics and economics. The authors do not seem to have fully incorporated the concept of noise into their understanding of statistical evidence. It can be challenging to explore a new continent without local guides who can show you the territory. That said, we appreciate the way the authors connect the statistical concept of variation to intuitions about cognition and decision-making, and we hope this book spurs further work in this direction.

Further Reading

Dallas, S.K., Liu, P.J., Ubel, P.L. 2019. Don't count calorie labeling out: calorie counts on the left side of menu items lead to lower calorie food choices. *Journal of Consumer Psychology*, 29, 60–69.

Francis, G., Thunell, E. 2020. Excess success in "Don't count calorie labeling out: calorie counts on the left side of menu items lead to lower calorie food choices." Meta-Psychology 4, MP.2019.2266.

Gelman, A. 2022. Pizzagate and "Nudge": An opportunity lost. *Statistical Modeling, Causal Inference and Social Science. https://statmodeling.stat.columbia.edu/2022/06/04/pizzagate-and-nudge-an-opportunity-lost.*

Kahneman, D. 2017. Comment on "reconstruction of a train wreck: how priming research went off the rails," by Ulrich Schimmack. *Replicability Index*.

Rusoff, J.W. 2022. Errors in economic predictions. *Think Advisor*.

Tosh, C., Greengard, P., Goodrich, B., Gelman, A., Hsu, D. 2024. The piranha problem: Large effects swimming in a small pond. *Notices of the American Mathematical Society*.

About the Authors

Gaurav Sood <waiting>

Andrew Gelman is a professor of statistics and political science and director of the applied statistics center at Columbia University. He has received many awards, including the Outstanding Statistical Application Award from the American Statistical Association and the award for best article published in the *American Political Science Review*. He has coauthored many books; his most recent is *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. He is editor of the "ethics and statistics" column.