Noise Pollution

Noise: A Flaw in Human Judgment
Daniel Kahneman, Olivier Sibony & Cass R. Sunstein (Random House 2021), 410 pages

Patrick Barry, rev'r[*]

   The authors of *Noise: A Flaw in Human Judgment* are a trio of intellectual heavy hitters: Nobel Prize-winner Daniel Kahneman, constitutional law scholar Cass Sunstein, and former McKinsey consultant (and current management professor) Olivier Sibony. As prolific as they are prominent, the three of them have collectively produced over fifty books and hundreds of articles, including some of the most cited research in social science.[1] If academic publishing ever becomes an Olympic sport, they'll be prime medal contenders, particularly if they get to compete as a team or on a relay. Their combined coverage of law, economics, psychology, medicine, education, finance, political science, corporate strategy, statistics, and even Star Wars gives the book the feel of a cognitive decathlon.[2]

   At the center of it all is a key distinction: the difference between *bias* and *noise*. Judgments are biased, the authors explain, when they are "systematically off target."[3] If, however, "people who are expected to agree

- -

[*] Clinical Assistant Professor, Director of Digital Academic Initiatives, University of Michigan Law School.

[1] The Google Scholar page for Kahneman credits his work with having received over 232,000 citations. *Daniel Kahneman*, Google Scholar, https://scholar.google.com/citations?user=E8z3WEYAAAAJ&hl=en (last visited May 19, 2022). And the one for Sunstein indicates a similarly large influence: 164,689 citations and counting. *Cass Sunstein*, Google Scholar, https://scholar.google.com/citations?user=ddq2_gkAAAAJ&hl=en (last visited May 19, 2022). Newer to the scholarly world, Sibony still comes in at a respectable 1,737 citations as of May 18, 2022. *Olivier Sibony*, Google Scholar, https://scholar.google.com/citations?user=PJARmj0AAAAJ&hl=en (last visited May 18, 2022).

[2] For a sense of the authors' cumulative range, *see, e.g.*, Bernard Garrette, Corey Phelps & Olivier Sibony, Cracked It! How to Solve Big Problems and Sell Solutions Like Top Strategy Consultants (2018); Daniel Kahneman, Thinking, Fast and Slow (2011); Olivier Sibony, You're About to Make a Terrible Mistake! How Biases Distort Decision-Making—and What You Can Do to Fight Them (2019); Cass Sunstein & Richard Thaler, Nudge: Improving Decisions about Health, Wealth, and Happiness (2009); Cass Sunstein, The World According to Star Wars (2016); Cass Sunstein, Simpler: The Future of Government (2013).

[3] Daniel Kahneman, Olivier Sibony & Cass R. Sunstein, Noise: A Flaw in Human Judgment 4 (2021).

end up at very different points around the target,"[4] then we have a different problem: the problem of noise.[5]

Failing to recognize and separate these two flaws in decisionmaking can have major consequences, especially given that

- trying to persuade a group of people who are biased—geographically, politically, economically, socially—is different than trying to persuade a group of people that is noisy;

- fixing an academic grading scheme that is biased is different than fixing an academic grading scheme that is noisy; and

- working through a set of feedback that is biased is different than working through a set of feedback that is noisy.

A major benefit of Kahneman, Sunstein, and Sibony's book is that it gives you a way to distinguish—and navigate—each of these situations.

I. Bias, noise, and dart boards

To help illustrate their bias vs. noise dichotomy, the authors begin the book with an example that involves a bullseye at a gun range.[6] When I summarize the main points of the example for my law students, however, I switch the visual to a bullseye on a dart board. I ask them to imagine that a group of people throw a bunch of darts. Each person aims directly for the bullseye. Each person tries their best. Yet when we take a look at where their darts end up, we notice that every single one of them lands slightly to the right of the bullseye. Not to the left. Not above. Not below. All cluster in the same spot to the right.

That's what bias is, according to Kahneman, Sibony, and Sunstein. The darts are, to return to the definition above, "systematically off target."

Think of the many studies that have uncovered racial bias and gender bias in the way hiring decisions are made,[7] criminal sentences are delivered,[8] and mortgage rates are offered.[9] There is a (depressingly) recognizable pattern to these forms of discrimination. We can predict how the next decision in the queue is going to go.

· ·

4 *Id.*

5 *Id.*

6 *Id.* at 3–5.

7 Lincoln Quillian, Devah Pager, Ole Hexel & Arnfinn H. Midtbøen, *Meta-analysis of Field Experiments Shows No Change in Racial Discrimination in Hiring Over Time*, 114 Proc. Nat'l Acad. Sci. 14 (2017).

8 Rhys Hester & Todd Hartman, *Conditional Race Disparities in Criminal Sentencing: A Test of the Liberation Hypothesis from a Non-Guidelines State*, 33 J. of Quantitative Criminology 77 (2017).

9 Justin Steil, Len Albright, Jacob Rugh & Douglas Massey, *The Social Structure of Mortgage Discrimination*, 33 Housing Stud. 759 (2018).

Or, to take a less grave example, consider a research paper by the economist Noland Kopkin called *The Nature of Regional Bias in Heisman Voting*.[10] Using a data set that stretched over twenty-five years, Kopkin found that the hundreds of journalists and other pundits who vote every year for college football's most prestigious award, the Heisman Trophy, have exhibited a consistent bias towards players from their own region.[11] Voters from the Northeast favor players from the Northeast. Voters from the Southwest favor players from the Southwest. And so on.

The bias isn't egregious, and Kopkin suggests that the overall effect is decreasing now that there are more and more ways to watch games from every region.[12] But if we imagine each of those votes as darts on the dart board we've been talking about, we'd probably see quite a bit of clustering. There'd be a cluster around the Northeast of the dartboard, representing the bias of voters from that region. There'd be a cluster around the Southwest of the dartboard, representing the bias of the voters from that region. There'd be clusters all over the place.

Not so with *noise*. When the problem is noise, there aren't any clusters. There aren't predictable patterns. There's simply a random assortment of darts.

II. Noisy judgments, major damage

Bias and noise are both big problems. But Kahneman, Sibony, and Sunstein worry that concerns about bias, however legitimate, have overshadowed concerns about noise. "The topic of bias has been discussed in thousands of scientific articles and dozens of popular books," they write, "few of which even mention the issue of noise."[13] Bias has become "the star of the show," while noise is treated as "a bit player, usually offstage."[14] Their book tries to correct that imbalance, a task they believe is particularly important given the stakes involved. Here are few of the areas they identify where noisy judgments can cause major damage:

· ·

10 Nolan Kopkin, *The Nature of Regional Bias in Heisman Voting*, 5 J. Sports Analytics 85 (2019). Kopkin has also found evidence of "own-race" bias. *See* Nolan Kopkin, *Evidence of Own-Race Bias in Heisman Trophy Voting*, 100 Soc. Sci. Q. 176 (Feb. 2019).

11 Each of the six regions—Northeast, Mid-Atlantic, South, Southwest, Midwest, and Far West—are given 145 votes. All living Heisman Trophy winners are also allowed to vote, and one collective vote is awarded based on a fan poll. Scott McDonald, *How the Heisman Trophy Winner is Selected, and When the Finalists are Named*, Newsweek (Dec. 22, 2020, 8:30 PM EST), https://www.newsweek.com/how-heisman-trophy-winner-selected-when-finalists-are-named-1556818#:~:text=Who%20are%20the%20Heisman%20voters,with%20145%20voters%20per%20region.

12 Kopkin, *supra* note 10, at 87.

13 Kahneman, *supra* note 3, at 6.

14 *Id.*

Doctor Diagnoses: "Faced with the same patient, different doctors make different judgments about whether patients have skin cancer, breast cancer, heart disease, tuberculosis, pneumonia, depression, and a host of other conditions."[15]

Child Custody Decisions: "Case managers in child protection agencies must assess whether children are at risk of abuse and, if so, whether to place them in foster care. The system is noisy, given that some managers are much more likely than others to send a child to foster care."[16]

Patent Applications: "The authors of a leading study on patent applications emphasize the noise involved: 'Whether the patent office grants or rejects a patent is significantly related to the happenstance of which examiner is assigned the application.'"[17]

III. Personality change

One source of these distortions is what the authors call *occasion noise*—when faced with the same decision at different times, people make conflicting judgments. Asked to review an identical set of X-rays several months apart, for example, a set of doctors disagreed with their original judgment between sixty-three percent and ninety-two percent of the time.[18] That's not doctors coming to a different conclusion than other doctors. That's doctors coming to a different conclusion than themselves.

Or consider a frequent criticism of personality tests like the Myers-Briggs Type Indicator. If you take the test more than once, there's a good chance you'll find out that your "personality" has changed.[19]

That happened to Adam Grant, an organizational psychologist at the University of Pennsylvania and author of bestselling books such as *Give and Take* and *Think Again*. In an article titled *Goodbye to the Myers-Briggs Typical Indicator, the Fad That Won't Die*, Grant shares the incompatible scores he received.[20] The first time he took the test he was classified as an "INTJ," meaning he was allegedly more introverted than extroverted, more intuiting than sensing, more thinking than feeling, and more judging

---

15 *Id.*

16 *Id.*

17 *Id.* at 7.

18 Robert Sutton, *How to Turn Down the Noise that Mars Our Decision-Making*, Wash. Post (May 21, 2021, 3:18 PM EDT), https://www.washingtonpost.com/outlook/how-to-turn-down-the-noise-that-mars-our-decision-making/2021/05/19/758be210-b370-11eb-9059-d8176b9e3798_story.html.

19 David J. Pittenger, *Measuring the MBTI . . . And Coming Up Short*, 54 J. Career Plan. & Emp. 48 (Nov. 1993); *see also* Joseph Stromberg & Estelle Caswell, *Why the Myers-Briggs Test is Totally Meaningless*, Vox (Oct. 8, 2015, 8:30 AM EDT), https://www.vox.com/2014/7/15/5881947/myers-briggs-personality-test-meaningless.

20 Adam Grant, *Goodbye to MBTI, the Fad That Won't Die*, Psych. Today (Sept. 18, 2013), https://www.psychologytoday.com/us/blog/give-and-take/201309/goodbye-mbti-the-fad-won-t-die.

than perceiving. These labels initially seemed to match his own image of himself. "Although I spend much of my time teaching and speaking on stage, I am more of an introvert—I've always preferred a good book to a wild party. And I have occasionally kept lists of my to-do lists."[21]

Yet when Grant took the same test a few months later, each of those classifications reversed. Now, apparently, he was a big-time extrovert. "Suddenly, I had become the life of the party, the guy who follows his heart and throws caution to the wind."[22]

Grant's experience is a textbook example of occasion noise and also one of the reasons he says that "when it comes to accuracy, if you put a horoscope on one end and a heart monitor on the other, the Myers-Briggs Test falls about halfway in between." [23] In other words, the test has a lot of noise and not much use.

IV. (Under) performance

The authors of *Noise* don't mention Grant's essay. But he is one of many academic luminaries who provides a cover blurb for the book. "Get ready," he raves, "for some of the world's greatest minds to help you rethink how you evaluate people, make decisions, and solve problems."[24] He has also done an extensive research project as a consultant for Facebook to help fix something the *Noise* authors devote an entire chapter to: employee performance reviews.[25]

One complaint about performance reviews—especially those that happen only once a year—is the time lag involved. The reviews come long after the person being reviewed could have used the instruction and guidance the process is designed to provide. Here's how a manager at PricewaterhouseCoopers, which is one of the many major companies that have moved away from annual performance reviews, expressed that frustration:[26] "You don't give elite athletes coaching at the end of the season. You give it in the middle of the game."[27]

. . . .

21 *Id.*

22 *Id.*

23 *Id.*

24 Kahneman, *supra* note 3.

25 Janelle Gale, Lori Goler & Adam Grant, *Let's Not Kill Performance Evaluations Yet*, Harvard Bus. Rev. (Nov. 2016), https://hbr.org/2016/11/lets-not-kill-performance-evaluations-yet.

26 Lillian Cunningham & Jena McGregor, *More U.S. Companies Moving Away from Traditional Performance Reviews*, Wash. Post (Aug. 17, 2015), https://www.washingtonpost.com/business/economy/more-us-companies-moving-away-from-traditional-performance-reviews/2015/08/17/d4e716d0-4508-11e5-846d-02792f854297_story.html.

27 Alexia Elejalde-Ruiz, *Companies are Scrapping Annual Performance Reviews for Real-Time Feedback*, Chi. Trib. (Apr. 22, 2016, 9:20 AM), https://www.chicagotribune.com/business/ct-performance-reviews-overhaul-0424-biz-20160421-story.html.

The authors of *Noise*, however, focus on a different problem. Discrepancies in evaluations often have more to do with who is doing the evaluating than with the employees themselves. Imagine that you ran a race and three different stopwatches evaluated how well you did compared to the other runners. One stopwatch said you finished second overall. Another said you finished eleventh. And the third didn't even put you in the top fifty.

Wouldn't that be kind of frustrating? Wouldn't you think something was wrong with the way your performance in the race was assessed?

Any student who has picked a class based on whether the teacher is a hard or easy grader has faced a similar issue. For over a century, research has shown that teachers vary widely on how they evaluate students.[28] In one of the most cited experiments, the same two English papers were given to 200 teachers. The authors of the study—Daniel Starch and Edward Elliott of the University of Wisconsin—were quite disturbed by the huge discrepancy in the grades the papers received. One paper, for example, earned a near perfect score from some teachers, but it received a failing score from others. "It is almost shocking to a mind of more than ordinary exactness," Starch and Elliot said of the overall results, "to find that the range of marks given by different teachers to the same paper may be as large as 35 or 40 points."[29]

When Starch and Elliot tried the same experiment with math teachers—a group presumably more committed to objective, stable standards—the variation persisted.[30] Identical student responses to questions about theorems, bisecting angles, and the hypotenuse of a triangle. Yet widely different grades. That's not bias. (There was no identifying information about the students' race, gender, or other characteristics which could have improperly influenced the teachers.) That, alarmingly, is noise.[31]

V. Decision hygiene

By the end of the book, it is hard not to think that we live in an exceedingly noisy world. There is noise in the way actuaries calculate

- -

28 For an overview of this research, including a discussion of a few studies that push back on the research that shows high grade variability, see Susan M. Brookhart et al., *A Century of Grading Research: Meaning and Value in the Most Common Educational Measure*, 86 Rev. Educ. Res. 803, 806–20 (2016).

29 Daniel Starch & Edward C. Elliott, *Reliability of the Grading of High-School Work in English*, 20 Sch. Rev. 442 (1912). For a more recent study, see Hunter M. Brimi, *Reliability of Grading High School Work in English*, 16 Prac. Assessment, Rsch. & Evaluation 1 (2011).

30 Starch & Elliott, *supra* note 29, at 254.

31 *Id.*

insurance premiums.[32] There is noise in the way judges decide asylum cases.[33] There is noise virtually everywhere, including in high-stakes judgments made every day in banks, start-ups, daycares, law firms, nonprofits, and the C-suites of Fortune 500 companies. It's enough to make you want to invest in a really good pair of earplugs.

A better approach, however, would be to follow the steps the authors suggest lead to good "decision hygiene."[34] The quotations below contain a few that one of those authors, Olivier Sibony, highlighted in an interview soon after the book was published.[35] I've then added some potential ways to apply them to the writing that lawyers and professors do.

Aggregate multiple independent judgments: "Whenever you have different people making judgments, rather than assign the judgment to one person or gathering three people to talk about it around the table, get them to make their judgments independently and take the average of that."[36] An appellate judge, for example, might canvas each of their clerks separately about a particularly hard case instead of—or at least before—holding a chambers-wide discussion about the issues involved. Group dynamics being what they are, you don't want one clerk's strong "Reverse" to prematurely influence (and perhaps even silence) another clerk's helpfully dissenting "Affirm."

Invest in competence: "Some people are going to be better than others at any judgment. In medicine, for instance, some diagnosticians are better than others. If you can pick the better people, that helps. The better people are going to be more accurate; they are going to be less biased but they're also going to be less noisy. There is going to be less random error in their judgments."[37] Recommendation letters are full of noise. How do you compare a candidate that one reference describes as "exceptional" with a candidate that a different reference describes as "amazing?"

One tactic is to evaluate the evaluators: Which recommenders consistently supply you with people who actually end up being well suited for the positions you are trying to fill? Many veteran judges, hiring partners, and admission officers already have informal networks of people and organizations that fulfill this "feeder" function. But if you're just starting out in one of these roles, it might be helpful to take a more systematic approach

---

32 Kahneman, *supra* note 3, at 23–33.

33 *Id.* at 6–7.

34 *Id.* at 226.

35 Olivier Sibony, *Sounding the Alarm on System Noise*, McKinsey Q. (May 18, 2021), https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/sounding-the-alarm-on-system-noise.

36 *Id.*

37 *Id.*

by keeping a tally of the success vs. dud ratio of your initial set of sources. You might also ask repeat recommenders to indicate how the current person they're touting stacks up against previous applicants they've sent your way. As the next tip from Sibony makes clear, comparison is key.

Use relative rather than absolute scales: "If you replace an absolute scale with a relative scale, you can eliminate a very big chunk of the noise. Think of performance evaluations again. Saying that someone is a 'two' or a 'four' on a performance-rating grid—even when you have the definition of what those ratings mean—remains fairly subjective, because what 'an outstanding performer' or 'a great relationship skill' means to you is not necessarily the same thing that it means to me. But if you ask, 'Are Julia's relationship skills better than those of Claudia?' that's a question I can answer if I know both Julia and Claudia. And my answers are probably going to be very similar to yours. Relative judgments tend to be less noisy than absolute ones."[38]

A helpfully visual way to operationalize relative judgments was suggested to me in graduate school in a class about pedagogy. Suppose, the teacher said, you are grading a bunch of papers. After you finish the first one, place it on the floor. Then move on to the next one. After you finish that one, place it on the floor as well—but be very deliberate about where it goes. If you think it's better than the first paper, it should go above that paper. If you think it's worse, it should go below.

Now repeat this same process with the rest of the papers, each time figuring out where precisely the most recent one fits among the set already ranked on the floor. Does it go above all but the top two? Below all but the bottom four?

You might even create large areas of physical space between key clusters. Perhaps the seventh, eighth, and ninth best papers are pretty similar in quality but each is significantly better than the tenth best paper. Or maybe there's a big drop off between number fifteen and number sixteen—the kind of gap that's less like the difference between a B+ and B and more like the difference between a B+ and a C-. Seeing two feet of flooring between those two papers (or exams, or resumes, or any other documents you're asked to evaluate) might helpfully separate them in your mental scoring system.

· ·

38 *Id.*

VI. Thankless but helpful

None of the decision hygiene ideas above are especially novel or sophisticated. Implementing them won't necessarily earn you any awards for innovative teaching or management. Nor will conducting the "Noise Audit" the authors attach as an appendix to the book.[39] As Sibony acknowledges, noise prevention is "a little bit thankless."[40]

But what you miss out in terms of gratitude and acclaim, you might gain in terms of efficiency, accuracy, and fairness. You don't need Daniel Kahneman's Nobel Prize in Economics to know that's a pretty good trade-off.

---

[39] Kahneman, *supra* note 3, at 23–33.

[40] Sibony, *supra* note 35.