

**LEBRIS**

We know  
books

**Also by DANIEL KAHNEMAN**

*Thinking, Fast and Slow*

**Also by OLIVIER SIBONY**

*You're About to Make a Terrible Mistake!: How Biases Distort  
Decision-Making—and What You Can Do to Fight Them*

**Also by CASS R. SUNSTEIN**

*Too Much Information: Understanding  
What You Don't Want to Know*

*Nudge: Improving Decisions About Health, Wealth, and Happiness*  
(with Richard H. Thaler)

# NOISE

*A Flaw in Human Judgment*

DANIEL KAHNEMAN  
OLIVIER SIBONY  
CASS R. SUNSTEIN



WILLIAM  
COLLINS

## Contents

<i>Introduction: Two Kinds of Error</i>	3
<b>Part I: Finding Noise</b>	11
1. Crime and Noisy Punishment	13
2. A Noisy System	23
3. Singular Decisions	34
<b>Part II: Your Mind Is a Measuring Instrument</b>	39
4. Matters of Judgment	43
5. Measuring Error	55
6. The Analysis of Noise	69
7. Occasion Noise	79
8. How Groups Amplify Noise	94
<b>Part III: Noise in Predictive Judgments</b>	107
9. Judgments and Models	111
10. Noiseless Rules	123

11. Objective Ignorance	137
12. The Valley of the Normal	148
<b>Part IV: How Noise Happens</b>	159
13. Heuristics, Biases, and Noise	161
14. The Matching Operation	176
15. Scales	187
16. Patterns	200
17. The Sources of Noise	210
<b>Part V: Improving Judgments</b>	221
18. Better Judges for Better Judgments	225
19. Debiasing and Decision Hygiene	236
20. Sequencing Information in Forensic Science	245
21. Selection and Aggregation in Forecasting	259
22. Guidelines in Medicine	273
23. Defining the Scale in Performance Ratings	287
24. Structure in Hiring	300
25. The Mediating Assessments Protocol	312
<b>Part VI: Optimal Noise</b>	325
26. The Costs of Noise Reduction	329
27. Dignity	339
28. Rules or Standards?	350
<i>Review and Conclusion: Taking Noise Seriously</i>	361
<i>Epilogue: A Less Noisy World</i>	377
<i>Appendix A: How to Conduct a Noise Audit</i>	379
<i>Appendix B: A Checklist for a Decision Observer</i>	387

<i>Appendix C: Correcting Predictions</i>	391
<i>Acknowledgments</i>	397
<i>Notes</i>	399
<i>Index</i>	439

LBRIS

We know  
books

NOISE

statements as they are or adapt them for any issues that matter to you, whether they involve health, safety, education, money, employment, entertainment, or something else. Understanding the problem of noise, and trying to solve it, is a work in progress and a collective endeavor. All of us have opportunities to contribute to this work. This book is written in the hope that we can seize those opportunities.

## PART I

# Finding Noise

It is not acceptable for similar people, convicted of the same offense, to end up with dramatically different sentences — say, five years in jail for one and probation for another. And yet in many places, something like that happens. To be sure, the criminal justice system is pervaded by bias as well. But our focus in chapter 1 is on noise — and in particular, on what happened when a famous judge drew attention to it, found it scandalous, and launched a crusade that in a sense changed the world (but not enough). Our tale involves the United States, but we are confident that similar stories can be (and will be) told about many other nations. In some of those nations, the problem of noise is likely to be even worse than it is in the United States. We use the example of sentencing in part to show that noise can produce great unfairness.

Criminal sentencing has especially high drama, but we are also concerned with the private sector, where the stakes can be large, too. To illustrate the point, we turn in chapter 2 to a large insurance company. There, underwriters have the task of setting insurance premiums for potential clients, and claims adjusters must judge the value of claims. You might predict that these tasks would be simple and

mechanical and that different professionals would come up with roughly the same amounts. We conducted a carefully designed experiment — a noise audit — to test that prediction. The results surprised us, but more importantly they astonished and dismayed the company's leadership. As we learned, the sheer volume of noise is costing the company a great deal of money. We use this example to show that noise can produce large economic losses.

Both of these examples involve studies of a large number of people making a large number of judgments. But many important judgments are *singular* rather than repeated: how to handle an apparently unique business opportunity, whether to launch a whole new product, how to deal with a pandemic, whether to hire someone who just doesn't meet the standard profile. Can noise be found in decisions about unique situations like these? It is tempting to think that it is absent there. After all, noise is unwanted variability, and how can you have variability with singular decisions? In chapter 3, we try to answer this question. The judgment that you make, even in a seemingly unique situation, is one in a cloud of possibilities. You will find a lot of noise there as well.

The theme that emerges from these three chapters can be summarized in one sentence, which will be a key theme of this book: *wherever there is judgment, there is noise — and more of it than you think*. Let's start to find out how much.

## CHAPTER 1

### Crime and Noisy Punishment

Suppose that someone has been convicted of a crime — shoplifting, possession of heroin, assault, or armed robbery. What is the sentence likely to be?

The answer should not depend on the particular judge to whom the case happens to be assigned, on whether it is hot or cold outside, or on whether a local sports team won the day before. It would be outrageous if three similar people, convicted of the same crime, received radically different penalties: probation for one, two years in jail for another, and ten years in jail for another. And yet that outrage can be found in many nations — not only in the distant past but also today.

All over the world, judges have long had a great deal of discretion in deciding on appropriate sentences. In many nations, experts have celebrated this discretion and have seen it as both just and humane. They have insisted that criminal sentences should be based on a host of factors involving not only the crime but also the defendant's character and circumstances. Individualized tailoring was the order of the day. If judges were constrained by rules, criminals would be treated in a dehumanized way; they would not be seen as unique

individuals entitled to draw attention to the details of their situation. The very idea of due process of law seemed, to many, to call for open-ended judicial discretion.

In the 1970s, the universal enthusiasm for judicial discretion started to collapse for one simple reason: startling evidence of noise. In 1973, a famous judge, Marvin Frankel, drew public attention to the problem. Before he became a judge, Frankel was a defender of freedom of speech and a passionate human rights advocate who helped found the Lawyers' Committee for Human Rights (an organization now known as Human Rights First).

Frankel could be fierce. And with respect to noise in the criminal justice system, he was outraged. Here is how he describes his motivation:

*If a federal bank robbery defendant was convicted, he or she could receive a maximum of 25 years. That meant anything from 0 to 25 years. And where the number was set, I soon realized, depended less on the case or the individual defendant than on the individual judge, i.e., on the views, predilections, and biases of the judge. So the same defendant in the same case could get widely different sentences depending on which judge got the case.*

Frankel did not provide any kind of statistical analysis to support his argument. But he did offer a series of powerful anecdotes, showing unjustified disparities in the treatment of similar people. Two men, neither of whom had a criminal record, were convicted for cashing counterfeit checks in the amounts of \$58.40 and \$35.20, respectively. The first man was sentenced to fifteen *years*, the second to 30 *days*. For embezzlement actions that were similar to one another, one man was sentenced to 117 *days* in prison, while another was sentenced to 20 *years*. Pointing to numerous cases of this kind, Frankel deplored what he called the “almost wholly unchecked and sweeping powers” of federal judges, resulting in “arbitrary cruelties

perpetrated daily,” which he deemed unacceptable in a “government of laws, not of men.”

Frankel called on Congress to end this “discrimination,” as he described those arbitrary cruelties. By that term, he mainly meant noise, in the form of inexplicable variations in sentencing. But he was also concerned about bias, in the form of racial and socioeconomic disparities. To combat both noise and bias, he urged that differences in treatment of criminal defendants should not be allowed unless the differences could be “justified by relevant tests capable of formulation and application with sufficient objectivity to ensure that the results will be more than the idiosyncratic ukases of particular officials, justices, or others.” (The term *idiosyncratic ukases* is a bit esoteric; by it, Frankel meant personal edicts.) Much more than that, Frankel argued for a reduction in noise through a “detailed profile or checklist of factors that would include, wherever possible, some form of numerical or other objective grading.”

Writing in the early 1970s, he did not go quite so far as to defend what he called “displacement of people by machines.” But startlingly, he came close. He believed that “the rule of law calls for a body of impersonal rules, applicable across the board, binding on judges as well as everyone else.” He explicitly argued for the use of “computers as an aid toward orderly thought in sentencing.” He also recommended the creation of a commission on sentencing.

Frankel’s book became one of the most influential in the entire history of criminal law — not only in the United States but also throughout the world. His work did suffer from a degree of informality. It was devastating but impressionistic. To test for the reality of noise, several people immediately followed up by exploring the level of noise in criminal sentencing.

An early large-scale study of this kind, chaired by Judge Frankel himself, took place in 1974. Fifty judges from various districts were asked to set sentences for defendants in hypothetical cases summarized in identical pre-sentence reports. The basic finding was that “absence of consensus was the norm” and that the variations across

punishments were “astounding.” A heroin dealer could be incarcerated for one to ten years, depending on the judge. Punishments for a bank robber ranged from five to eighteen years in prison. The study found that in an extortion case, sentences varied from a whopping twenty years imprisonment and a \$65,000 fine to a mere three years imprisonment and no fine. Most startling of all, in sixteen of twenty cases, there was no unanimity on whether any incarceration was appropriate.

This study was followed by a series of others, all of which found similarly shocking levels of noise. In 1977, for example, William Austin and Thomas Williams conducted a survey of forty-seven judges, asking them to respond to the same five cases, each involving low-level offenses. All the descriptions of the cases included summaries of the information used by judges in actual sentencing, such as the charge, the testimony, the previous criminal record (if any), social background, and evidence relating to character. The key finding was “substantial disparity.” In a case involving burglary, for example, the recommended sentences ranged from five years in prison to a mere thirty days (alongside a fine of \$100). In a case involving possession of marijuana, some judges recommended prison terms; others recommended probation.

A much larger study, conducted in 1981, involved 208 federal judges who were exposed to the same sixteen hypothetical cases. Its central findings were stunning:

*In only 3 of the 16 cases was there a unanimous agreement to impose a prison term. Even where most judges agreed that a prison term was appropriate, there was a substantial variation in the lengths of prison terms recommended. In one fraud case in which the mean prison term was 8.5 years, the longest term was life in prison. In another case the mean prison term was 1.1 years, yet the longest prison term recommended was 15 years.*

As revealing as they are, these studies, which involve tightly controlled experiments, almost certainly understate the magnitude of noise in the real world of criminal justice. Real-life judges are exposed

to far more information than what the study participants received in the carefully specified vignettes of these experiments. Some of this additional information is relevant, of course, but there is also ample evidence that irrelevant information, in the form of small and seemingly random factors, can produce major differences in outcomes. For example, judges have been found more likely to grant parole at the beginning of the day or after a food break than immediately before such a break. If judges are hungry, they are tougher.

A study of thousands of juvenile court decisions found that when the local football team loses a game on the weekend, the judges make harsher decisions on the Monday (and, to a lesser extent, for the rest of the week). Black defendants disproportionately bear the brunt of that increased harshness. A different study looked at 1.5 million judicial decisions over three decades and similarly found that judges are more severe on days that follow a loss by the local city’s football team than they are on days that follow a win.

A study of six million decisions made by judges in France over twelve years found that defendants are given more leniency on their birthday. (The defendant’s birthday, that is; we suspect that judges might be more lenient on their own birthdays as well, but as far as we know, that hypothesis has not been tested.) Even something as irrelevant as outside temperature can influence judges. A review of 207,000 immigration court decisions over four years found a significant effect of daily temperature variations: when it is hot outside, people are less likely to get asylum. If you are suffering political persecution in your home country and want asylum elsewhere, you should hope and maybe even pray that your hearing falls on a cool day.

### Reducing Noise in Sentencing

In the 1970s, Frankel’s arguments, and the empirical findings supporting them, came to the attention of Edward M. Kennedy, brother of the slain president John F. Kennedy, and one of the most influential

members of the US Senate. Kennedy was shocked and appalled. As early as 1975, he introduced sentencing reform legislation; it didn't go anywhere. But Kennedy was relentless. Pointing to the evidence, he continued to press for the enactment of that legislation, year after year. In 1984, he succeeded. Responding to the evidence of unjustified variability, Congress enacted the Sentencing Reform Act of 1984.

The new law was intended to reduce noise in the system by reducing "the unfettered discretion the law confers on those judges and parole authorities responsible for imposing and implementing the sentences." In particular, members of Congress referred to "unjustifiably wide" sentencing disparity, specifically citing findings that in the New York area, punishments for identical actual cases could range from three years to twenty years of imprisonment. Just as Judge Frankel had recommended, the law created the US Sentencing Commission, whose principal job was clear: to issue sentencing guidelines that were meant to be mandatory and that would establish a restricted range for criminal sentences.

In the following year, the commission established those guidelines, which were generally based on average sentences for similar crimes in an analysis of ten thousand actual cases. Supreme Court Justice Stephen Breyer, who was heavily involved in the process, defended the use of past practice by pointing to the intractable disagreement within the commission: "Why didn't the Commission sit down and really go and rationalize this thing and not just take history? The short answer to that is: we couldn't. We couldn't because there are such good arguments all over the place pointing in opposite directions.... Try listing all the crimes that there are in rank order of punishable merit.... Then collect results from your friends and see if they all match. I will tell you they won't."

Under the guidelines, judges have to consider two factors to establish sentences: the crime and the defendant's criminal history. Crimes are assigned one of forty-three "offense levels," depending on their seriousness. The defendant's criminal history refers principally

to the number and severity of a defendant's previous convictions. Once the crime and the criminal history are put together, the guidelines offer a relatively narrow range of sentencing, with the top of the range authorized to exceed the bottom by the greater of six months or 25%. Judges are permitted to depart from the range altogether by reference to what they see as aggravating or mitigating circumstances, but departures must be justified to an appellate court.

Even though the guidelines are mandatory, they are not entirely rigid. They do not go nearly as far as Judge Frankel wanted. They offer judges significant room to maneuver. Nonetheless, several studies, using a variety of methods and focused on a range of historical periods, reach the same conclusion: the guidelines cut the noise. More technically, they "reduced the net variation in sentence attributable to the happenstance of the identity of the sentencing judge."

The most elaborate study came from the commission itself. It compared sentences in bank robbery, cocaine distribution, heroin distribution, and bank embezzlement cases in 1985 (before the guidelines went into effect) with the sentences imposed between January 19, 1989, and September 30, 1990. Offenders were matched with respect to the factors deemed relevant to sentencing under the guidelines. For every offense, variations across judges were much smaller in the later period, after the Sentencing Reform Act had been implemented.

According to another study, the expected difference in sentence length between judges was 17%, or 4.9 months, in 1986 and 1987. That number fell to 11%, or 3.9 months, between 1988 and 1993. An independent study covering different periods found similar success in reducing interjudge disparities, which were defined as the differences in average sentences among judges with similar caseloads.

Despite these findings, the guidelines ran into a firestorm of criticism. Some people, including many judges, thought that some sentences were too severe — a point about bias, not noise. For our purposes, a much more interesting objection, which came from numerous judges, was that guidelines were deeply unfair because