

Daniel Kahneman, Olivier Sibony
and Cass R. Sunstein

**Noise: A Flaw in Human
Judgment**

Made by Blinkist



These key insights in blinks were written by a team of experts at Blinkist. We screen the world of nonfiction to choose the very best books. Then, we read them deeply and transform them into this concise format that brings you the most inspiring ideas from the text.

Maybe these blinks will inspire you to dig deeper, or maybe they're enough to start you thinking and then on to something new. However you read blinks, we hope they help you become an even brighter you.

What's in it for me? A deep dive into the mysterious world of noise.

Imagine you're holding a stopwatch in your hand. Without looking, start the watch and then stop it after exactly ten seconds.

If you do this a couple of times in a row you'll notice that hitting ten seconds on the dot is just about impossible.

Sometimes you'll be a little short, sometimes you go long. Sometimes you're off by milliseconds. Other times you're off by a whole second, or even more. Either way, through this little experiment, you'll end up with a set of errors that have no discernible pattern and no apparent cause.

This is an example of *noise*, or random mistakes in judgment. And while your errors in this little stopwatch experiment are innocent enough, as you're about to learn, variations in judgment like these

can have far more serious consequences.
Welcome to the strange world of noise.

In these blinks, you'll learn

- what the weather has to do with your chances of getting into college;
- why you – and everyone else – are terrible at predicting the future; and
- why our narrative-seeking brains wreak havoc on our judgments.

Unrelated and unpredictable factors can have an alarming impact on human judgment.

To get a better grasp on the random and strange nature of the kind of noise we're talking about here, let's imagine you're a high school senior, and you and your best friend are self-professed academic nerds. You've both earned straight As, nailed the SATs, and landed admissions interviews at the same Ivy League university.

You go to your interview and everything goes swimmingly. Your high marks impress the admissions officer and you cross the campus back to your car feeling great, the sun on your face and a cool breeze at your back.

Your friend has her appointment with the same admissions officer on the following day. Just like for you, her interview is a smooth ride. But when she

leaves, the rain clouds that have gathered all afternoon break open into a downpour.

Weeks pass, and you each receive a letter from the admissions office. Turns out, they've rejected you but accepted your friend. Your mind reels. Why? What does she have that you don't?

Here's the first key message: Unrelated and unpredictable factors can have an alarming impact on human judgment.

According to a 2003 paper evocatively titled "Clouds Make Nerds Look Good" by behavioral scientist Uri Simonsohn, the weather might have made the difference. Simonsohn discovered that on cloudier days, college admissions officers pay more attention to grades and scores.

On sunnier days, on the other hand, admissions officers are more sensitive to nonacademic qualities, meaning that on

the day of your interview, the officer might have been more interested in athletics and artistic talent than straight As and SAT scores.

Then again, perhaps the admissions officer's decision had nothing to do with the weather at all, and more to do with the interviewees that preceded you. Perhaps those students were great candidates, and the admission officer just didn't want to go on an acceptance streak.

But wait. Other irrelevant factors may also have influenced the decision. The admissions officer might have been hungry; he may have felt that sunny day was too hot, despite the air conditioning in the office; his hometown football team might have just lost an important game. Researchers have shown that each of these irrelevant factors can affect the decisions of bank loan officers, baseball umpires, physicians, and judges.

Importantly, in all of these scenarios, one person repeatedly confronts substantially the same situation, yet makes different judgments. Researchers call this variability *occasion noise*, and it's one of the major categories of noise. But it's not the only one.

Noise and bias are different, though bias can lead to noise.

Let's do another thought experiment, one that takes place at a carnival. Specifically, at a shooting arcade.

You and a friend, BB rifles in hand, have just fired several metal pellets at paper targets hanging at the far end of the range. You're both terrible shots, but in different ways.

On your paper target, the misses are scattershot. Taken at arm's length, you can see there's no pattern. Your shots are noisy.

But your friend's paper target tells a different story. His shots cluster together, but not on the bullseye. His shots tend to be low and to the left. It's as though he believes the real bullseye is actually down there. Or perhaps the

barrel of his rifle is bent. What's behind his systematic inaccuracy?

The key message here is: Noise and bias are different, though bias can lead to noise.

Whenever we make errors systematically, we call that *bias*.

Day to day, we employ the term to describe prejudice for or against certain groups of people. In the field of psychology, the term is often used to identify cognitive mechanisms that skew our judgments.

Take *conclusion bias*, for example, which causes us to bend our judgments toward a desired outcome, making us interpret evidence in a skewed way. Just consider one Miami immigration court, where the chance of getting asylum was found to swing from 5 to 88 percent depending on which of two judges heard the case. These two judges were most likely

suffering from conclusion bias. Needless to say, this kind of bias can have life-changing consequences.

If asylum decisions were pellets in a BB gun, the two Miami judges would each produce paper targets with shots clustered off target. But if the asylum decisions of the entire Miami courthouse were mapped out, including the variable decisions of other judges, the courthouse's paper target would show a scattershot mess.

This kind of variability, where judgments within a system are unjustifiably inconsistent from one another, is called *system noise*.

Remember the occasion noise of the Ivy League admissions officer? That too may have resulted from bias. But whether we're sussing out occasion noise or system noise, we have to examine our paper target at arm's length. If we hold

the target too close to our eyes, the noise ceases to be apparent.

Now, let's turn to another area prone to noise: predictions.

When we make predictions we're easily led astray by what feels right.

A judge deciding bail faces a mighty responsibility. Should she keep a defendant in jail pending trial, or let him go?

If she wrongly denies bail, the defendant will lose his freedom and perhaps his job. His family might even lose their home. None of this loss will have served justice. On the other hand, if she wrongly grants bail, he may flee and worse, commit another crime.

Weighing these consequences, a bail judge calls upon her experience and the record before her to predict what the defendant would do if released.

Unfortunately, humans – and that includes judges – are terrible at making accurate predictions.

The key message here is: When we make predictions we're easily led astray by what feels right.

In 2018, a team led by computation and behavior researcher Sendhil Mullainathan created an algorithm that produced bail judgments. They fed it the results of approximately 760,000 real-life bail hearings and found that the algorithm would have simultaneously lowered the jail population and crime committed by released defendants by 24 to 42 percent.

Other studies have found that a rudimentary formula that considers only two factors – the defendant's age and the number of court dates they've missed – also outperforms human judges.

This begs the question: Why would experts with years of training and experience fall a distant third to

algorithms and back-of-the-envelope math?

The answer is simple. Judges are human.

When we attempt to predict the future, we're seeking closure. It's an attempt to solve a mental puzzle, and when we come up with an answer, we experience an internal signal that says, *Yes, that's it!*

A satisfying prediction coheres to the way we see the world, and the power of this emotional reward often blinds us to a basic limitation of predictions: *objective ignorance.*

We don't know what we don't know, and what we do know might be wrong, incomplete, or misleading.

Rules and algorithms are also ignorant. More so, in fact. But they're free from internal signals, perceptions of the world, and emotional rewards. In short, they

outperform us because they're free from noise.

We ignore noise because it doesn't make for a good story.

You might have noticed that blinks often begin with a story.

We sketch out a time and place, an event that involves a character, someone with a goal and obstacles to overcome. We do this because the human mind loves a story, and information that works within a story tends to stick.

So far we've looked at noise and some of its consequences in the justice and college admissions systems. As we've said, anywhere people make judgments, there's bound to be noise. But if noise is everywhere, why haven't we heard more about it?

The key message here is: We ignore noise because it doesn't make for a good story.

Much of the psychological insight of recent decades has explored our deep attachment to narrative. The human mind, we've learned, understands the world by making stories to explain what we observe.

For example, psychologists have identified what's called the *fundamental attribution error*, our habit of crediting or blaming people for outcomes that are better explained by mere circumstance. In other words, we see characters and plots everywhere. When our reality is challenged, the psychological mechanism of *naive realism*, the self-reinforcing belief that we perceive reality just as it is, bolsters our narratives by ruling out troublesome counter-narratives. And when the truly unexpected occurs, the mind works to bring it within what the authors call the *valley of the normal*, where the strange is made understandable by attributing a cause in hindsight.

That brings us to the main point here: noise resists narrative. Noise isn't causal and it doesn't cohere to our patterns of understanding. To the degree that it's a story at all, it's a frustrating and apparently meaningless one.

Without a story to accommodate noise, it escapes our notice. We either miss it entirely, edit it out of our awareness, or perceive it as an instance of bias.

Bias, after all, does well in a story. It has causal force.

Noise, on the other hand, can only be observed statistically. The variability of bail denials, college admissions, asylum hearings, and hiring decisions may feel like bias when experienced firsthand. And as we've said, bias may be a contributing factor in some cases. But when we back up and view these phenomena in the aggregate, their random, chaotic nature becomes apparent.

We can cancel out noise by averaging multiple, independent judgments on a single question.

Before moving on, let's just quickly recap what we've learned so far, which is this: Wherever there's human judgment, there'll be noise and, sometimes, that noise can have alarming, life-changing consequences.

So, that naturally leads to the question: What can we do about it? How do we turn down the noise?

To begin answering this question, let's go back to the fall of 1906 when Francis Galton, a polymath and cousin to famed evolutionist Charles Darwin, visited a county fair in Plymouth. While strolling through the stalls, he came across an ox-weighing competition. Galton, who among other things, was a theorist of intelligence, listened with curiosity as close to 800 villagers gave their best

estimates of the ox's weight. No one got the correct answer: 1,198 pounds. When the competition was over, Galton asked the organizers to lend him the tickets for statistical analysis.

When plotted on a graph, the estimates were scattershot. Over here and under there by variable, unpredictable amounts. The villagers were noisy. But, when Galton calculated the mean of the villagers' estimates, he noticed something surprising. It was nearly perfect, only off by a single pound.

The key message here is: We can cancel out noise by averaging multiple, independent judgments on a single question.

Galton had stumbled across a phenomenon now known as the *wisdom-of-crowds effect*. In hundreds of different circumstances, getting independent judgments from multiple judges and then averaging their

responses has been shown to bring you close to the truth.

When you ask people to guess the number of jelly beans in a jar, the distance of a randomly chosen city, or the temperature a week from today, their answers will vary. They'll be noisy. However, when averaged out, the noise in one response counteracts the contradictory noise in another. The noise cancels itself out.

But the wisdom of the crowd comes with some essential caveats.

First, each judge must be independent of the others. When you ask a group a question all together, the individuals respond as much to the group as they do to the question itself.

Also, the wisdom of the crowd only bears out when each individual considers the exact same case. Asking each person

a different question will get you nowhere.

Finally, the wisdom of the crowd doesn't guard against bias. If the crowd shares a bias, a systematic error in judgment, the mean of the members' responses will only distill and amplify that bias. For example, if a hiring committee suffers from bias against women, that bias will only appear more salient when we aggregate and average its judgments on a female job candidate.

To fight noise, you've got to first make it visible through a noise audit.

So far, we've spent a fair bit of time talking about judges and the random and occasionally inexplicable variability in the sentences they hand out. The random quality of this injustice didn't escape the notice of Marvin E. Frankel, a US District Judge.

Early in his career, Frankel realized that he could, for example, send a convicted bank robber to prison for up to 25 years, or he could choose a sentence of one day. The judgment, Frankel saw, ultimately depended on his own views, predilections, and biases.

In 1973 Frankel published a book showcasing disparities in sentencing for substantially similar crimes. One case in point was a small-time check counterfeiter who got 30 days, and

another that got 15 years for essentially the same crime.

Individual anecdotes, however, can be explained away. So, Frankel set out to create a more durable, systematic portrait.

The key message here is: To fight noise, you've got to first make it visible through a *noise audit*.

In 1981, Frankel led a research team that asked 208 federal judges to sentence criminal defendants from 16 fictional vignettes.

Frankel's team presented the scenarios to each judge individually and then mapped out the variation between the judges' proposed sentences in each case. The study, and several others like it, statistically proved that shocking variability, not consistency, was the norm in criminal sentencing.

Frankel showed that an audit can determine the level of noise in any institution where a pool of experts takes on substantially similar cases. Here's how it's done.

First, determine your bullseye. Decide how much variability in judgment is acceptable. For an insurance executive, the question might be, what's a tolerable difference in the payout recommended by separate claims adjusters inspecting the same flooded basement?

Next, gather your judges, or claims adjusters as the case may be, and present them with scenarios. Make sure to provide them with a numerical expression for each judgment, such as years in a prison sentence or dollars in an insurance payout.

Finally, map out the judges' answers in relation to your bullseye. And voilà, you now have a diagnostic portrait of the

noise in your institution. Now, what can you do about it?

Decision hygiene can reduce noise, and like regular hygiene, it's about discipline and prevention.

Imagine you're on an operating table, about to go under the knife. Just before you nod off under the influence of a general anesthetic, the lead surgeon steps over to the sink. She lathers her hands up with soap and then places them under the hot water. With that one simple act, she has prevented an unknowable number of pathogens from entering your body.

We can do much the same with noise. By adopting principles we'll call *decision hygiene*, we can make our world substantially less noisy.

The key message here is: Decision hygiene can reduce noise, and like regular hygiene, it's about discipline and prevention.

The first step to developing decision hygiene – the equivalent to the surgeon washing her hands with soap – is learning how to stop and think statistically before making any important decision.

We learned in a previous blink that our narrative-seeking minds make a story of everything. We dive in and imbue the particulars of a case with cause and meaning. While this is a natural tendency, it's an invitation for noise. Instead, you should aim to take what Kahneman calls *the outside view*: frame every case in reference to a body of other similar cases.

For example, let's say you've got a new CEO at work and you're wondering whether she'll be successful. The CEO's education, reputation, and performance history might give you some clues, but it's a mass of complex and potentially misleading information.

A less noisy approach would begin by exploring outcomes from similar situations. For example, you could find out the average rate of turnover of CEOs in your industry, or explore how often new CEOs result in rising share values.

While building a strong statistical framework, you should resist premature intuition.

We all love a judgment that feels right, but you want it to feel right for the right reasons. Stow that gut feeling you may have about the CEO's alma mater or that hunch on what went down at her last job. Instead, save the emotional reward for a judgment that jibes with a well-founded view of what's most likely.

On a related note, creating a judgment that's both complex and coherent is emotionally rewarding, but that reward can mislead you. If possible, break tough cases into separate questions and hand them off to independent judges.

For example, connecting the dots between your CEO's tenure and the company's stock value might be a fun mental puzzle, but it could also be irrelevant.

We have one more critical principle of decision hygiene, and it comes in the form of a cautionary tale.

For noise reduction to stick, it's essential that judges buy in.

In 1984, Judge Frankel was victorious. Congress enacted the Sentencing Reform Act and soon after instituted strict sentencing guidelines based on an analysis of 10,000 real cases. Under the new rules, judges could only consider the crime and the defendant's criminal history. The judge would assign a numerical value to each and the resulting score dictated the range of possible sentences.

As a result, noise in sentencing plummeted.

For example, before the act, a man convicted of dealing drugs might have faced a variability of years in his sentence depending purely on the judge that happened to hear the case. Afterward, that variability went down to a range of a few months.

Judges across the country complained bitterly. They'd fine-tuned their sense of justice through years of study and experience, and now their discretion was gone, replaced by a crude math problem.

The key message here is: For noise reduction to stick, it's essential that judges buy in.

In 2005, the Supreme Court struck down the Sentencing Reform Act for technical reasons. A few years later, Harvard law professor Crystal Yang conducted an analysis of 4,000 criminal cases sentenced after the act's demise. The disparity between harsh sentences and the average had doubled. Personal values had emerged as a basis for sentencing. Noise had returned.

In hindsight, Judge Frankel and his allies missed a crucial step in their campaign for noise reduction. They'd failed to make the judges agree on the ultimate purpose of judgment.

The goal of judgment should be accuracy, not personal expression.

Unlike literary criticism, competitive sport, filmmaking or any other field where diversity in opinion and style breeds richness and growth, variability among experts judging substantially similar cases is a problem. When two radiologists independently read the same x-ray and arrive at different conclusions, one of them is wrong. In other words, when judges step up to the shooting arcade, they must first agree on the same bullseye.

Once the judges agree that accuracy is the highest priority, the auditors should invite them to create the test scenarios. Failing to do so assures the audit will meet with hostile scrutiny. Next, the judges have to see the extent and costs of the noise.

Kahneman, for example, conducted an audit in an insurance company and

discovered that underwriters averaged a 55 percent difference in the premiums they set for customers. For the underwriters to appreciate the importance of reducing this noise, they had to understand that losses from over- and underpricing ran in the hundreds of millions of dollars.

Finally, when instituting decision hygiene, judges must take part in formulating practical, system-specific rules that balance the reducing noise with other costs. For example, a backlash against rising crime led some US states to adopt a “three strikes” rule that mandated life imprisonment for defendants convicted of three felonies. The rule reduced noise, but did so without accounting for a defendant’s history, the severity of her crime, or her capacity for rehabilitation.

Audits, hygiene, rules, habit, and prevention. Noise reduction isn’t very glamorous work. But by now, you’ll likely

agree that noise exacts a terrible cost. It wastes resources, spurs injustice, and results in personal tragedy. It erodes faith in institutions of law, medicine, education, and work. Now that we've uncovered it, it's our job to reduce it.

Final summary

The key message in these blinks:

Random and unwanted variability in human judgment is everywhere, and whether we see it or not, we pay a heavy price. The good news is we can reduce noise if we change our mindset and adopt principles of prevention.

Actionable advice:

Tap into the *wisdom of the crowd within*. As you learned in blink five, averaging multiple, independent judgments of a single question can counteract the noise of those judgments and leave you with a strikingly accurate answer. Thing is, if you ask yourself the same question multiple times, you can achieve the same effect.

Try it. Over the next few days, ask yourself the following question: *what*

share of the world's airports is in the United States?

The average of your responses will be relatively noiseless and surprisingly close to the truth.

Spoiler alert: The answer is 32 percent.

Nice work! You're all done with this one.

We publish new books every week at
blinkist.com.

Come and see – there's so much more to learn.

Inspired to read the full book?

[Get it here.](#)

Copyright © 2014 by Blinks Labs GmbH.
All rights reserved.