In press: *The American Journal of Psychology*
Running head: Noise

Twilight of Human Judgment

Review of 'Noise: A flaw in human judgment'
by Daniel Kahneman, Olivier Sibony, and Cass R. Sunstein

Joachim I. Krueger
Brown University
2,283 words of text
(excluding references)

Correspondence:

Joachim I. Krueger
Department of Cognitive, Linguistic & Psychological Sciences
Brown University
190 Thayer St.
Providence, RI 02912
Phone: (401) 863-2503
Home page: http://research.brown.edu/research/profile.php?id=10378

Keywords: noise, error, bias, judgment, prediction, rationality

Note: This book is published by Little, Brown Spark, New York, 2021.
ISBN 978-0-316-45140-6, list price: $32, 454 pages, hardcover.

The river is 5 feet deep, on average.
– Last words of a 6-ft-tall man before drowning, paraphrased after Howard Marks

With *Noise*, Kahneman, Sibony, and Sunstein (hereafter: KSS) have produced a book of practical relevance for business organizations, the law, and the rest of us. This book is not, as I first feared it might be, another list of biases, fallacies and other disasters of human judgment. Instead, there is a compelling inner logic, moving from analysis and diagnosis to intervention and application. With this structure, *Noise* has the feel of an engineering text or a manual for practice. The first two sections are rather atheoretical. Indeed, how can one have a good theory of randomness if noise is what is left after all the systematic, and theoretically explicable, components of variation have been removed? Whether anything can even be truly random is a question on which mathematicians disagree (Prömel, 2005). KSS's approach is to move methodically through the mechanics of psychometric analysis and variance partitioning to isolate noise, that is, randomness we are unable to explain. They then show how, even if it cannot be explained, this residue can be tamed with the aid of audits, rules, and repeated judgments. *Et voilà*, less noise and better judgments!

The narrative becomes livelier in the third section, where KSS illustrate the dangers of noise pollution with examples from the legal environment and the world of business strategy. A psychological factor of concern is the widespread underestimation of the magnitude of noise and the common resistance to reducing it with the cold fingers of science and bureaucracy (Moore, 2020). KSS take these kinds of resistance seriously and acknowledge their partial validity. They do not propose a grand resolution, which may not be achievable on principle anyway. Human judgment, with its biases and errors, will continue to be with us, but perhaps it can be civilized. So, in the end, *Noise* is a significant book because it goes beyond diagnosing what is deplorable in human judgment; it also offers guidelines for what to do about it. In this review, I highlight some of *Noise's* main points and then explore some of the adaptive uses of noise, which KSS acknowledge but do not emphasize.

The shooter takes aim

To set the stage, KSS discuss target practice. A shooter aims at a target and produces a cloud of bullet holes around the bull's eye with a slight overall displacement to the right. The bull's eye is the true value, *T*; the difference between the average of the holes and T is the bias, *b*; and the remaining dispersion is the error, *e*, or noise proper. The holes – or 'judgments' – can be modeled as $J = T + b + e$, and the mean squared error, MSE, between *J* and *T* can be decomposed into the MSE of *b* and the MSE of *e*. To make this scenario one-dimensional and numerical, suppose $T = 10$ and the judgments are 8, 10, 12, 14, and 16. The total MSE = 12, which is the sum of MSE $b = 4$ and MSE $e = 8$. Averaging the individual estimates eliminates noise and brings the bias into focus. Accuracy is improved even if the distribution of *J* does not bracket *T*. If absolute differences between *J* and *T* are used, accuracy will not improve without bracketing, but it will not deteriorate either (Larrick, Mannes, & Soll, 2012). Hence, when there is a *T* that is in principle knowable, there is no excuse for not producing multiple aggregable judgments. Many human judgments are one-time affairs, but the wisdom of the crowd has taught us that even individual judges can benefit from estimating a quantity or an intensity repeatedly and then taking the average of their own judgments (Herzog & Hertwig, 2009; Krueger & Chen, 2014). This works if *T* is not only an algebraic element in the modeling of *J,* but also a causal force that pulls *J* toward itself. In other words, aggregated judgments are better than individual judgments as long as at least one of the individual judgments contains a bit of valid information.

Most statistical modeling of this sort begins with the assumption of independence, which means that what KSS call *occasion noise* is a random distribution of *e*, where no *e* can be predicted from any other, not even from the one that just preceded it. The assumption of independence is often violated in the real world. Judges remember their last estimates, and they might either double down and repeat them or deliberately produce estimates that are vastly different from the earlier ones. In the former case, these judges might be praised for their noiseless efforts, or they might be faulted for tumbling down their own judgmental cascade (Krueger & Massey, 2008). In the latter case, judges

might be faulted for the extra noise they are creating, or they might be lauded for creating partially independent estimates, which help the discovery of *T* after aggregation. Only an 'audit,' preferably one with access to truly independent estimates or *T* itself, can settle this issue. The deliberate reduction or amplification of noise may thus be beneficial or detrimental depending on circumstances. To know which it is, additional efforts are required.

The loss of *T*

The true value *T* in any judgment task plays a paradoxical role in psychological science. If *T* is known, we can model judgment and study the components of its lack in accuracy. In the shooter vignette, we know where the bull's eye is; we just want to see how well we can hit it. Here, noise is the expression of a trembling hand. In estimation tasks, noise is the expression of a trembling mind. If *T* is unknown, the battle against noise can proceed as before, but the size of the bias remains unknown. Noise reduction yields improvements, but no indication of how much improvement there is relative to the remaining error due to bias. KSS note that often we face *objective ignorance*, as there are limits to what can be known.

When T is unknown or unknowable, human values often fill the void. Reviewing evidence from medicine, organizations, and the law, KSS document dramatic, even scandalous, volumes of noise in the judgments made by presumed *respect experts*, that is, credentialed individuals who are supposed be able to make accurate and reliable judgments. Again, however, these judges' overconfidence, illusions of validity, and overblown respect paid to them by an awed audience obscure many of their inaccuracies. KSS apply their methods of noise reduction in this context as well. Judges themselves can be modeled statistically, as Paul Meehl, Lewis Goldberg, and Robyn Dawes had shown a series of classic publications. Simple, and even improper, linear models perform better than the judges and their trembling minds. Judgments can be audited for noise with the help of other judges or observers, and with the use rankings and comparisons with other, even hypothetical, cases. Beyond simple spreadsheet models loom algorithms capable of deep learning and pattern detection. These algorithms, if nothing else, raise the specter of a twilight of human judgment. Yet,

KSS reassure readers not to worry too much. Accurately forecasting the limits of such algorithms is itself a hard problem. One is tempted to consult a panel of experts and aggregate their predictions.

The loss of *T* is particularly apparent, and troublesome, in the domain of judicial decision-making. What is the proper sentence for second-degree murder? KSS note the horrendous between- and even within-judge noise in such decisions and the ongoing battle over binding sentencing requirements. When *T* is missing, the issue is not accuracy but procedural justice and fairness. The Achilles heel of the noise-reduction argument is that great reliability can come with great injustice. Religiously inspired justice may be highly consistent but draconian. Would one want to live in a society where a convicted murderer gets an unpredictable sentence of either a 5-year prison term or a beheading, or would one prefer a society where every convicted murderer is beheaded? When estimates of *T* are averaged, shrinking noise indicates growing accuracy. However, when diverse value judgments are averaged, a political dimension asserts itself. When a majority view is invested with power, the dissenters have grounds to complain that their values are being spurned and their interests ignored. In an earlier book, Sunstein (2005) discussed the value of dissent for a healthy society. When dissent reduces to noise, however, it runs the risk not only of not being heard, but of also admitting adverse material consequences. KSS are aware of these dangers, but being squarely focused on the war on noise, their exploration of these issues is rather tentative.

To make the case for the reduction of judgment noise by mechanical and statistical means is to vote for a greater bureaucratization of life. It is a commitment to one horn of Weber's dilemma. Weber (1968/1922) saw bureaucracy as a key to modernity. At its best, a bureaucracy provides a set of rational processes and applies them fairly. At its worst, a bureaucracy subjects a population to a yoke of arbitrary demands. But even at its best, a bureaucracy creates what Weber called an 'iron cage.' It limits, by design, the human freedom to err.

I will not attempt to defend the freedom to err on scientific grounds because I do not know how to separate the facts from the values, but I will point to some heuristic advantages. The first argument we have already seen. Often, we need to create more estimates, accept the errors they bring,

and then take advantage of averaging. Merely repeating the first estimate yields a false consistency. If Dom bets on heads and I bet on tails when flipping a coin, one of us must be wrong. However, it is also clear that one of us will win. If we both bet on heads, we are either both wrong, or we are forced to split the purse. The second argument is foraging, an activity nature demands of her creatures so they can survive when food gets scarce in the familiar patch. To find new fertile ground, animals have to engage in noisy exploration (Stephens, Brown, & Ydenberg, 2007). For the third argument, I lean on Messrs. Dostoevsky and Cockcroft. Dostoevsky repeatedly – although exact references escape me – expressed a reactant attitude (Brehm, 1966) to behaving predictably. Even if, he asserted, it could be shown that behavior X yields the greatest rewards, he would choose behavior Y. To Dostoevsky, the escape from being predictable was the refuge of freedom. George Cockcroft, under the pen name Luke Rinehart (1971), wrote *The Diceman*, a tale of a man who did what the dice told him. The Diceman's contribution was to select options and to assign a probability to each; after that, he rolled the dice and obeyed their verdict (Krueger, 2010). Bizarre as this strategy may sound, it makes a provocative point. Cockcroft satirized the psychotherapeutic fads of his day, where sufferers were assured that they carried true values, desires, motives and capabilities within themselves, and that their job was to discover and act on them. They had to find out who they truly were and what they truly wanted. They were, in other words, challenged to discover their *T*s. Cockcroft thought this was oppressive nonsense (for a recent critique of the deep mind hypothesis, see Chater, 2018, reviewed by Massaro, 2019). Cockcroft proposed that a richer life can be lived if multiple *T*s are brought forth and acted upon. What looks like error, on this view, is spicy variety.

Three random remarks

Let us close with three critical notes on this generally impressive book. First, KSS favor rankings over absolute ratings, mainly because rankings force the judge to take an outside view by making inter-stimuli comparisons. The use of rankings abandons the standard psychometric equation of $J = T + b + e$. Unreliability may survive in the form of low correlations between sets of rankings, but the errors associated with individual ranks are no longer independent. Any information about the

underlying range of scores is lost. If re-test reliability coefficients are low, one suspects a restricted range. In the performance evaluation domain, for example, rankings ensure that 20% of the cases will fall into the lowest quintile irrespective of the overall level of accomplishment. Being suppressed by a rigid bureaucratic algorithm, these cases (i.e., people) have cause to complain.

Second, the call for repeated judgments and inter-stimuli comparisons sounds like a tribute to repeated-measures research designs. It is thus not without irony that Kahneman and colleagues have for decades favored single-shot, between-subjects, research designs to document the poverty of human judgment. Would the story of human cognition be more hopeful than the one told in *Thinking, fast and slow* (Kahneman, 2011) if the human subjects had been given better opportunities to de-noise themselves (cf. Schulze & Hertwig, 2021)?

Third, what may be said for human thinking once the mechanical processes of error reduction have been deployed? If averaging is not thinking because a dead mechanism can achieve the result, and if noise is irreducible randomness, what is left? KSS offer no theory of cognition. In their hands, human thought is rather like the god of the gaps in apophatic theology: a master of a shrinking domain. The question of what mechanical models are missing has been asked for generations. Roger Penrose (1994), for example, argued that Gödel's incompleteness theorem entailed that human minds can never be fully modeled as machines. It may be disappointing to note that something is missing without being able to show what it is that is missing. It is even more disappointing, however, to pretend that the story is complete.

References

Brehm, J. W. (1966). *A theory of psychological reactance.* Academic Press.

Chater, N. (2018). *The Mind Is Flat: The remarkable shallowness of the improvising brain*. Yale University Press.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20*, 231-237.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus & Giroux.

Krueger, J. I. (2010). Beyond free will and determinism: Take a chance with the Dice Man. *Psychology Today Online*. https://www.psychologytoday.com/us/blog/one-among-many/201009/beyond-free-will-and-determinism

Krueger, J. I., & Chen, L. J. (2014). The first cut is the deepest: Effects of social projection and dialectical bootstrapping on judgmental accuracy. *Social Cognition, 32*, 315-335.

Krueger, J. I., & Massey, A. L. (2009). A rational reconstruction of misbehavior. *Social Cognition, 27*, 785-810.

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In J. I. Krueger (Ed.), *Social judgment and decision making* (pp. 227-242). Psychology Press.

Massaro, D. (2019). The richness of flatness. *American Journal of Psychology, 132*, 378-382.

Moore, D. A. (2020). *Perfectly confident*. HarperCollins.

Penrose, R. (1994). *Shadows of the mind: A search for the missing science of consciousness*. Oxford University Press.

Prömel, H. J. (2005). Complete disorder is impossible: The mathematical work of Walter Deuber. *Combinatorics, Probability and Computing*, *14*, 3–16.

Rinehart, L. (1971). *The Diceman*. Overlook Press.

Schulze, C., & Hertwig, R. (2021). A description–experience gap in statistical intuitions: Of smart babies, risk-savvy chimps, intuitive statisticians, and stupid grown-ups. *Cognition, 210*. https://doi.org/10.1016/j.cognition.2020.104580

Stephens, D.W., Brown, J.S., & Ydenberg, R.C. (2007). *Foraging: Behavior and Ecology*. University of Chicago Press.

Sunstein, C. R. (2005). *Why societies need dissent*. Harvard University Press.

Weber, M. (1968/1922). *Economy and society*. Bedminster Press. Originally published in German.