Noise: A Flaw in Human Judgment. [Book Review]

Michael Cohen
Old Dominion University, mcohe006@odu.edu

KAHNEMAN, Daniel; SIBONY, Olivier; SUNSTEIN, Cass R.

Noise: A Flaw in Human Judgment.

New York: Hachette Book Group, 2021. ISBN 978-0-316-45140-6.

Michael Cohen[a]

Many are already familiar with the past work of Daniel Kahneman, including his previous book *Thinking Fast and Slow* co-authored with his long-time colleague Amos Tversky. The most recent work by Kahneman et al. is even more impactful in terms of its implications for public decision-making. While the previous work revealed widespread biases present in judgment, the current work constitutes an unrelenting attack on our collective conception of judgment itself. While established institutions can readily accept the notion that improvements in formalized processes are needed, it is quite another matter to ask them to reconsider the entire basis for how they conduct business. Presented with evidence indicating that their past strategic decisions may have been no more insightful than a coin toss, it would be understandable for leaders to ignore such findings altogether. Such disregard would be an understandable, and yet potentially very costly, mistake to make.

The authors find "noise", or unexplained variability in decisions, in a host of places where judgment is routinely called for. In some examples provided, such as the case of a particular insurance underwriting firm, the variability between decisions was so large that the authors compared it to a de facto lottery (p. 24). So, how do we make sense out of a world where our expectations of unanimity in judgment are regularly falsified by our experiences? In short, we choose not to see contrary evidence and continue nurturing the illusion of agreement (pp. 30-31). As the authors point out, we find it easy to maintain this illusion due to a common professional language, common rules, and a common understanding of what factors should influence our decisions. We assume that everyone sees the world the same way that we do.

These findings should set off alarm bells for military professionals in the West. Reliance upon professional judgment is enshrined in the North Atlantic Treaty Organization's (NATO) Operational Planning Process (OPP), among a myriad of other processes. As an example, a commander is expected to exercise professional judgment in selecting a course of action from among those prepared by the unit planning staff. Is what we imagine to be a reasonable choice made by a professional based on their education and experience in fact just another lottery? Could the staff just skip the step of briefing the commander and flip a coin instead? While this may or may not lead to an increase in correct choices,

[a] Strome College of Business (School of Public Service), Old Dominion University. Norfolk, Virginia, USA. Email: mcohe006@odu.edu.

replacing the commander with an algorithm has at least a reasonable chance of doing so (pp. 112-138) While a commander could make a different decision based on such factors as their mood, weather, stress, fatigue, time of day, and based simply on the order of options presented (pp. 86–90), an algorithm will not.

At this point one could interject and argue that not only is the commander a professional, but they are also an expert in the purposeful and orderly application of violence. Clearly, expertise is an advantage both over the algorithm as well as the random coin toss, right? The value of expertise as a quality varies across disciplines, however. The authors acknowledge that true experts exist in domains where their skills can be verified and compared with the results. Chess masters fall easily into this category. There are also "respect-experts" (p. 226), whose expertise is entirely social and reputational. Wine tasters fall into this category. So, where is our notional commander on this spectrum between a chess master and a wine taster? The uncomfortable truth is the commander could be either. If a country is at peace, there are few ways to assess a commander's real aptitude in their chosen profession. Aside from being continuously involved in one war or another, such as modern Russia that nearly has been like this since its establishment in 1991,[1] there is simply no way to fully assess performance in combat.

"Most of the time professionals have confidence in their own judgment. They expect that colleagues would agree with them, and they never find out whether they actually do. In most fields, a judgment may never be evaluated against a true value and at most will be subjected to vetting by another professional who is considered a respect-expert. Only occasionally will professionals be faced with a surprising disagreement, and when it happens, they will generally find reasons to view it as an isolated case." (p. 369)

There exist some means of distinguishing between the judgments of various respect-experts, however, and the authors tell us that the general mental ability (GMA, formerly referred to as IQ, or Intelligence Quotient) is one of the most potent among them (pp. 228-232). This is the type of intelligence measured by a standardized test, and not the sort of intelligence that a selection panel discerns from someone's speech, dress, mannerisms, social circle, photograph, performance evaluations (more on that from the authors later), or the university which they graduated from. In other words, centralized military promotion selection boards for officers in the United States (and perhaps other countries) select for every type of measure of intelligence except for the type that most accurately correlates with better judgment. Also correlated with better judgment, according to the authors, is active open-minded thinking (p. 234). In a world that values standardization, discipline, and conformity, we can expect this to be largely lacking.

Performance evaluations, now generally assumed to be a proxy for tests of judgment based on their widespread usage in personnel decisions, are also found by the authors to be noisy. As it turns out, ratings of personnel not only vary among different supervisors, but the judgments of an individual supervisor can vary based on the noise factors discussed above by the authors (weather, time of day, order of options, stress level, fatigue, whether the local sports team won the night before, etc.). Not only might our notional unit staff achieve as good a judgment (or better) in selecting between options

---

[1] Since its inception in 1991, the Russian Federation has been continuously involved in foreign and domestic military conflicts, with the only possible exceptions being periods between 1997-1999 and 2009-2014.

by a coin toss rather than appealing to the commander, but the commander also may have been selected by the equivalent of a coin toss. Forced ranking systems (limiting the number of favourable assessments that can be allocated for a rated population) are also noisy, the authors find. The implementation of such systems often introduces more problems than the systems solve. It's not possible for 98% of managers to be in the top 20%, the authors admit (p. 294), but it is possible that your team consists disproportionately of high or low performers. The performance of small teams does not perfectly correlate to a normal distribution curve. The result: noise in judgments.

All is not lost, for the authors propose some methods to introduce "decision hygiene" into judgment processes to improve their quality. The Mediating Assessments Protocol developed by the authors (p. 323) combines many of these into one single process for implementation. Some of these include appropriately sequencing information, structuring a decision into independent assessments, grounding in an outside view of the problem, and aggregating the individual assessments of participants. The superiority of process over judgment and intuition is accepted by the authors, who simultaneously acknowledge that it may receive less welcome among leaders and managers trained to believe otherwise. In many ways, this process seems similar to the steps of the Operational Planning Process, at least until we come to the time for the commander to exercise sole, unfettered judgment in selecting between the courses of action proposed by the staff. Perhaps if that last step entailed more dialectical interaction between the commander and the staff (akin to the estimate-talk-estimate process proposed by the authors on p. 322) a better and less noisy decision might be reached. Why go through a structured process only to flip a coin at the end?

Another means of reducing noise discussed in the work is the substitution of rigid rules for vague standards. Instinctively, military professionals will recoil from any suggestion that combat decisions could be relegated to rules. These professionals would probably suggest substituting intuition and judgment in their place, but we have already received a sufficient warning from the text about going down that path. The criticism that rigid rules could fail us in fluid and unique circumstances is not without merit, however. According to the researchers, deciding between rules and standards is often a decision as to which type of costs in terms of errors one is the most keen to avoid: those costs caused by variance in judgment, or the costs of errors introduced by rigid rules (p. 357).

Kahneman et al. remind us that, while it may be uncomfortable for us to imagine that our judgment is often inferior to even the simplest of linear models, this is by now a well-established fact (p. 367). We might therefore consider using models and algorithms not only in designing courses of action, but also perhaps in deciding between them. In light of the arguments and evidence provided by the authors, it may even be time to reconsider the aversion in the West to Russia's style of tactical staff planning, with its heavy reliance on mathematical nomograms and selection from a menu of tried-and-true tactical combinations. Such methods may not only yield quicker tactical decisions with less staff deliberation,[2] but also perhaps better decisions as well.

---

[2] Grau, L. and Bartles, C. (2016). The Russian Way of War: Force Structure, Tactics, and Modernization of The Russian Ground Forces. Fort Leavenworth: Foreign Military Studies Office, pp. 51-59.

This work by Kahneman et al. should be a required reading both for military professionals, and for civilians who work in the defence sector. The authors raise the uncomfortable point that while we might be willing to acknowledge and correct bias in our judgments, we are far less willing (or even unwilling) to admit to the variance among judges or between judgments. Some of the noise-reducing strategies raised in the work (some of which have been referenced above) could go far in helping us to arrive at better solutions to tactical and operational problems. They could even help us select and promote better military professionals for positions requiring good judgment. It would be a far less expensive investment than turning wine tasters into chess masters in our next engagement with a peer-level adversary.

Note: The ideas and opinions expressed in this piece are those of the author, and in no way convey or express the official view of either the United States Government, Department of Defense [sic], or U.S. Army.