

Balancing the evidential scales for the mental unconscious¹

Aliya R. Dewey

Centre for Philosophy & AI Research,
Friedrich-Alexander-Universität Erlangen-Nürnberg

Review of **Ben R. Newell & David R. Shanks**, *Open Minded: Searching for Truth about the Unconscious Mind*, MIT Press, 2023, 234 pp., \$45.00, ISBN 9780262546195.

In *Open Minded: Searching for Truth about the Unconscious Mind*, Ben R. Newell & David R. Shanks (henceforth, N&S) challenge the popular claim that much of human judgment and decision-making is explained by unconscious processes. A tremendous amount of evidence has been offered in support of this claim, so N&S dedicate Part I of the book to arguing that (a) much of this evidence fails to replicate and (b) the little evidence that does replicate is better explained without appeals to unconscious processes. Their approach to the literature is pessimistic, which some readers (like myself) will find refreshing, but which the authors seem to expect will make some readers uncomfortable. Thus, N&S dedicate Part II of the book to arguing that their pessimistic stance is justified by the fact that “the scientific ecosystem... is skewed in ways that lead to the generation of vast swathes of junk science” (p. 129). They point to issues like p-hacking, HARKing (hypothesising after results are known), and publication bias and they call for solutions like strict pre-registration measures, changes to the review system, and more computational modelling.

The book is thought provoking and has valuable insights but suffers from unclear inferences, puzzling organisation, and lack of engagement with the philosophical literature. These issues emerge immediately in Chapter 1, where N&S set up the terms of the debate. First, they endorse a reportability conception of consciousness, which is a form of access consciousness (a la Block, 1995): roughly, a process is conscious if and only if that process is accessible to subjective report (a la Dohaene & Naccache, 2001). Second, the authors note that there is a weak sense of unconsciousness, such that it is just the absence of “reportability consciousness”. Of course, this trivialises the claim that judgment and decision-making is caused by unreportable processes, since they are caused by events like action potentials in individual neurons that are obviously unreportable. So, they introduce a more robust conception of unconsciousness: roughly, a process is unconscious in the robust sense if and only if that process is *mental* yet inaccessible to subjective report. Of course, this raises the question: what could make an unconscious process mental? Surprisingly, N&S offer no answer at all, and so fail to clearly specify what view they are challenging. This makes the rest of their argument rather difficult to evaluate.

Throughout the book, N&S negatively compare the robust sense of unconsciousness to the notion of black matter (e.g., p. 44, p. 53), suggesting that they don’t think it is plausible.

¹ This is a penultimate draft. Please cite the version that is forthcoming in *Philosophical Psychology*.

If unconsciousness isn't a plausible notion on theoretical grounds, though, then we can conclude *a priori* that it won't feature in the best explanation of *any* empirical evidence. There would be no need for their careful consideration of empirical evidence throughout the rest of the book. But N&S do spend half the book reviewing empirical evidence, suggesting that they do have a plausible robust conception of unconsciousness in mind that they intend to reject on empirical (rather than theoretical) grounds. I just wish N&S clearly articulated the position that they mean to be rejecting. This seems negligent given that there are multiple cogent conceptions of the mental unconscious in the philosophical literature (e.g., Searle, 1992; Hassin, 2013; Berger, 2014; Phillips & Block, 2016; Krickel, forthcoming).

N&S dedicate the rest of Part I to arguing that (a) most evidence for unreportable processes fails to replicate and (b) the little evidence that does replicate is better explained by appealing to reportable processes only. Since they fail to provide a plausible robust conception of unconsciousness, though, the rest of Part I falls a bit flat: they really only show that the replicable evidence *can* be explained by appealing to reportable processes only (let's call these "full-reportability" explanations), without offering any reasons to believe that these explanations are better than alternative explanations that appeal to a coherent sense of unconscious processing (let's call these "partial-reportability" explanations). This is disappointing, to be sure, but it is no trivial task for the authors to show that full-reportability theories explain all replicable evidence of judgment and decision-making—making it a live empirical option to reject partial-reportability explanations.

In Chapter 2 and later in Chapter 6, N&S develop full-reportability explanations for delayed feelings of agency on the Libet paradigm, pointing behaviour in patients with blindsight, and confabulation effects. In particular, they argue that subjects can fail to report mental processes because the prompting for reporting isn't sufficiently specific (as in the Libet paradigm), subjects are biased against responding (as in blindsight), or subjects are initially able to report mental processes but reportability is susceptible to rapid memory decay or distortion (as in confabulation effects)—rather than because those mental processes are intrinsically unreportable. This is a clever strategy. They seem to be suggesting that a mental process is unreportable in the relevant sense if and only if it is *intrinsically* unreportable—if and only if there are no extrinsic, auxiliary conditions that make it reportable. Obviously, the evidence that a mental process isn't reportable will always fall short of this: the observed non-reportability of a mental process can always be attributed to extrinsic, auxiliary conditions. Therefore, there is always room to look for empirical evidence of extrinsic, auxiliary conditions (e.g., specific reporting, liberal response bias) that do make apparently unreportable processes accessible to subjective report.

N&S are surprisingly effective at exploiting this opportunity and finding empirical evidence that processes that initially appear to be unreportable often prove to be reportable under different extrinsic auxiliary conditions. This goes a long way towards challenging the received view that so much of human judgment and decision-making is caused by unconscious mental processes. More importantly, it raises our evidential standards for appeals to unconscious processing: empirical evidence doesn't justify attributions of unconsciousness to mental processes unless we can show that they are unreportable *in favourable conditions for subjective reporting*.² This point

² There is a potential problem here. If blindsight is an unfavourable condition for subjective reporting in the relevant sense, then there are no favourable conditions for subjective reporting in blindsight patients and it is impossible to

is my main takeaway from the book and hence, I think these are the two best chapters in the book. A reader pressed for time and interested only in the core claim of the book might do well to read just these two chapters.

In Chapters 3–5, N&S challenge evidence of priming effects and the popular type of explanation that they are the result of unconscious biases. N&S note how many priming effects have failed to replicate, and they exploit the same strategy to argue that the priming effects that do replicate are caused by mental processes that are reportable in favourable auxiliary conditions. However, a puzzling feature of these chapters is the emphasis on rationality: using strategies pioneered by Gerd Gigerenzer and colleagues (e.g., Gigerenzer et al., 2000), N&S argue that priming effects can be construed as rational, even though they have frequently been construed as irrational. I struggle to see how this is relevant to the issue of unconscious processing (e.g., for an account that construes priming as an unconscious and rational, see Mandelbaum, 2015). In fact, M N&S say that “judgments and preferences can’t reflect unconscious biases if they’re not even biases” (p. 78) but the issue under debate is whether judgments and decisions reflect unconscious preferences, not whether those preferences count as biases. Without a clear connection between rationality and consciousness or mentality, these three chapters read like an interesting but prolonged tangent from the main argument.

In Chapter 7, N&S criticise the overuse of dichotomies in psychology—most famously in dual-process theories. The relevant dichotomy for the book’s purpose is between consciousness and unconsciousness, but their chapter mostly focuses on other dichotomies, e.g., between intuition and reflection. They point to studies by Stephens et al. (2018, 2019) that show that most dissociations become statistically insignificant if we replace implausible linearity assumptions with more plausible monotonicity assumptions during statistical analysis (for a response to this analysis, see Dewey, 2023). Again, the focus on dissociation studies is puzzling because dissociation studies are rarely used to find evidence for unconscious processing: N&S don’t cite any evidence that purports to dissociate reportable and unreportable processes. Once again, then, the relevance of this chapter to the main argument is unclear.

Overall, Part I deploys some clever strategies that make full-reportability explanations of judgment and decision-making a live empirical option. The effectiveness of these strategies indicates that it’s more difficult than expected to show that mental processes are unreportable in the relevant sense. In fact, it’s more difficult to show that mental processes are unreportable across *all* extrinsic, auxiliary conditions than it is to show that mental processes are reportable in at least one extrinsic, auxiliary condition. This gives full-reportability explanations an unfair evidential advantage over partial-reportability explanations: they can be confirmed by a single piece of evidence, whereas partial-reportability explanations can only be confirmed by many convergent pieces of evidence. As a result, Part I left me with many questions. How do we fairly arbitrate between full- and partial-reportability explanations now that we know that empirical evidence for unconscious processing is harder to come by than evidence for conscious

justify attributions of either consciousness or unconsciousness to blind sight patients. The same problem will be present for consciousness in any other health condition that affects reportability. This seems to be a difficult problem, but I don’t think it’s reason to reject the favourability proviso that N&S seem to advocate for reportability accounts of unconsciousness. Rather, I think it’s better treated as an open problem that warrants constructive philosophical attention. I thank Reviewer 1 for raising this issue.

processing? How should psychologists shoulder the increased evidential burden for finding any unreportable mental causes of behaviour? How do we individuate the favourable conditions for reportability, so that we can effectively and efficiently adjudicate whether a mental cause of behaviour is reportable in the relevant sense? Are there theoretical reasons to prefer a conception of mentality that is so closely linked to reportability that all (or, at least, most) mental causes are reportable?

Rather than answering any of these pressing questions, though, N&S dedicate Part II to answering what I found to be a much less pressing question: “How did we become hoodwinked into believing that our unconscious mind has a hold on behaviour?” (p. 119). In response, they discuss academic fraud and p-hacking (Chapter 8), publication bias, researcher degrees of freedom, HARKing (hypothesising after results are known), low sample sizes, and dubious relationships between sample and effect sizes (Chapter 9), and the incentives that reinforce bad scientific methodology and limit the uptake of better alternatives like mandatory pre-registration (Chapter 10). The discussion is persuasive (albeit familiar for readers who follow these issues) and makes effective use of several examples (like money priming), but it is barely integrated with Part I. N&S might have done better to publish Part II as a separate book. Perhaps, they expected Part I to shake the reader’s confidence in the current process of psychology and they expected Part II would be appropriate to validate this experience, arguing that the current process of psychology doesn’t warrant the reader’s confidence. That was not my response, though, so Part II fell a bit flat: although it’s inherently interesting, it answered none of the pressing questions raised by Part I. Other readers may have a similar experience.

N&S only revisit the specific topic of unconsciousness in Chapter 11, when they call for better theories of judgment and decision-making. They (rightfully) complain that partial-reportability explanations have been theoretically imprecise in their appeals to unconscious mental causes. They press the issue of *construct validity* against partial-reportability explanations in particular: there are many ways to measure unconscious processes, but there’s often very little evidence that different measures are measures of the same unconscious process. However, they fail to register that construct validity is inherently more difficult for unreportable than reportable processes: language offers a uniquely precise way of directing reports to the same mental process, but it’s unavailable for unreportable mental processes. Once again, then, this makes evidence of unreportable processes less accessible and this needs to be accounted for in a fair comparison of full- vs. partial-reportability explanations. The authors don’t register this evidential inequality, so they end up treating the absence of evidence as the evidence of absence, despite their own numerous injunctions throughout the book against this fallacy.

Overall, then, *Open Minded* provides an accessible and refreshingly sceptical whirlwind tour for readers who are new to the psychology of judgment and decision-making. For those who are already familiar with this literature, it is a thought provoking and occasionally rewarding but ultimately frustrating read. I would only recommend Chapters 2 and 6 to readers who are pressed for time and specifically interested in the relationship between empirical evidence and theories of the unconscious mind.

References

- Berger, J. (2014). Mental processes, conscious and nonconscious. *Philosophy Compass*, 9(6): 392–401. <https://doi.org/10.1111/phc3.12140>
- Block, Ned (1995). On a confusion about a function of consciousness. *Brain & Behavioral Sciences*, 18(2): 227–47.
- Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79(1–2): 1–37. [https://doi.org/10.1016/s0010-0277\(00\)00123-2](https://doi.org/10.1016/s0010-0277(00)00123-2).
- Dewey, A. R. (2023). Metacognitive control in single- vs. dual-process theory. *Thinking & Reasoning*, 29(2), 177–212. <https://doi.org/10.1080/13546783.2022.2047106>
- Hassin, R. R. (2013). Yes it can: On the functional abilities of the human unconscious. *Perspectives on Psychological Science*, 8(2): 195–207. <https://doi.org/10.1177/1745691612460684>.
- Krickel, Beate (2023). The unconscious mind worry: A mechanistic-explanatory strategy. *Philosophy of Science*, 90(1): 39–59.
- Mandelbaum, E. (2015). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, 50(3): 629–58.
- Phillips, I. (2018). Unconscious perception reconsidered. *Analytic Philosophy*, 59(4): 471–514. <https://doi.org/10.1111/phib.12135>.
- Phillips, I. & Block, N. (2016). Debate on unconscious perception. In B. Nanay (ed.), *Current Controversies in Philosophy of Perception* (pp. 165–92). London: Routledge.
- Searle, J. R. (1992). *The rediscovery of the mind*. MIT Press. [https://doi.org/10.1016/0004-3702\(95\)90037-3](https://doi.org/10.1016/0004-3702(95)90037-3)
- Stephens, R. G., Dunn, J. C., & Hayes, B. K. (2018). Are there two processes in reasoning? The dimensionality of inductive and deductive inferences. *Psychological Review*, 125(2), 218–244. <https://doi.org/10.1037/rev0000088>
- Stephens, R. G., Matzke, D., & Hayes, B. K. (2019). Disappearing dissociations in experimental psychology: Using process-trace analysis to test for multiple processes. *Journal of Mathematical Psychology*, 90: 3–22. <https://doi.org/10.1016/j.jmp.2018.11.003>