Optimality is critical when it comes to testing computation-level hypotheses

Laura S. Geurts [1], Andrey Chetverikov [1,#], Ruben S. van Bergen [1,#], Ying J. Zhou [1], Andrea Bertana [1] & Janneke F.M. Jehee [1,*]

[1] Donders Institute for Brain, Cognition and Behavior; Radboud University Nijmegen,

* Corresponding author (janneke.jehee@donders.ru.nl)

[#] these authors contributed equally

ABSTRACT

We disagree with Rahnev and Denison that optimality should be abandoned altogether. Rather, we argue that adopting a normative approach enables researchers to test hypotheses about the brain's computational goals, avoids just-so explanations, and offers insights into function that are simply inaccessible to the alternatives proposed by Rahnev and Denison.

While we concur with Rahnev and Denison (2018) that proving the optimality of human behavior should never be a goal in and of itself, we disagree that optimality should be discarded altogether. Rather, we argue that an ideal observer approach is the only valid approach for understanding the nature of the computational problems that the brain is trying to solve (Marr, 1982).

Understanding the brain's computational goals is key to understanding human behavior. Knowing what the system is trying to do, and why, can inform hypotheses about the algorithms that the brain uses to achieve these goals, and how these algorithms are implemented in neurons and connectivity (see Marr, 1982). To test computational theories, researchers typically translate its hypotheses into predictions that can be compared against human data. While this translation can take many forms, we argue that the best recipe for converting computational goals into predictions is provided by the normative approach, because this approach correctly specifies what kind of behavior participants ought to display under the hypothesized goals. For any other description, the translation from hypothesis into prediction may be imprecise, making it impossible to determine whether the hypothesis or rather its translation was incorrect, when predicted behavior fails to match human data.

Consider, for example, visual orientation perception. Human orientation perception tends to be biased towards vertical and horizontal (cardinal) orientations, such that a tilted (oblique) stimulus appears slightly more cardinal than it really is (Tomassini, Morgan, & Solomon, 2010). In a seminal study, Girshick and colleagues (2011) hypothesized that such biases arise because observers, when judging stimulus orientation, use knowledge about the distribution of orientations in the natural environment (where cardinal orientations are predominant). To address this hypothesis, the authors presented human participants with random orientation stimuli, asking them to estimate the orientations they had seen. Girshick et al. (2011) then specified the behavior of an ideal observer who combines noisy sensory measurements of a stimulus with knowledge about its prior probability. By using this approach to model human behavior, the authors were able to characterize the prior beliefs about orientation that the participants applied in their perceptual estimates. Interestingly, this prior distribution resembled the actual distribution of orientations in the environment, corroborating the theoretical predictions.

This study nicely illustrates how the ideal observer framework can be exploited to not only describe, but also explain behavior, by starting with a specific hypothesis about what the system is trying to do (minimizing perceptual error in a natural environment by taking into account natural orientation statistics), and translating this hypothesis into a computational model that predicts behavior. Precisely because human behavior matched that of the ideal observer, rather than some arbitrary formulation, the findings provided strong evidence for the computational theory, and offered insights into function that are simply inaccessible to the non-normative alternatives (e.g., 'bag of tricks' or 'neural network models', p. 54) proposed by Rahnev and Denison.

Note that in the example discussed above, human behavior was, in fact, suboptimal for the experimental situation in which the participants were tested: when presented with a uniform distribution of orientation stimuli, it would make little sense for an observer to show biased judgments towards vertical or horizontal orientations. Yet, the ideal observer framework enabled the researchers to address what kinds of knowledge the observers did bring to the task, by comparing human task behavior against that of an ideal observer using prior knowledge of the orientation statistics of the natural environment.

We argue that it is precisely these kinds of situations that are informative. The ideal observer framework provides a powerful tool to implement hypotheses, and as such, offers insight into the knowledge and

goals that a human observer brings to the task at hand. When human behavior does not conform to the theory's predictions, one should rather conclude that the theory makes incorrect assumptions about the computational goals of the system and try to improve the theoretical assumptions, instead of refuting the normative tools that merely implemented the theory.

REFERENCES

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926-932.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman and Company.

Rahnev, D., & Denison, R. N. (2018). Suboptimality in Perceptual Decision Making. *Behavioral and Brain Sciences*, 1–107.

Tomassini, A., Morgan, M.J., & Solomon, J.A. (2010). Orientation uncertainty reduces perceived obliquity. *Vision Research*, 50, 541-547.