

Biological constraints on interpretable theories of human decision-making

Science eLetters, July 20, 2021, <https://www.science.org/doi/10.1126/science.abe2629#elettersSection>

Kenway Louie (*Research Assistant Professor, New York University*)

Hayley Brooks (*PhD Student, University of Denver*)

Peter Sokol-Hessner (*Assistant Professor, University of Denver*)

In their recent paper (1), Peterson et al. present a promising new approach examining how humans make decisions by combining large-scale choice data with machine learning algorithms. This approach capitalizes on the function approximation power of deep neural networks to examine psychologically constrained - and thus interpretable - models.

While conceptually powerful, this approach sidesteps a different, fundamental constraint on human decision-making: biological constraints. Biological limitations have proven critical in sensory systems, where efficient coding strategies directly shape neural and perceptual responses (2-4), principles which may also apply to valuation and choice (5, 6). For example, unlike models operating on option attributes (1), neural systems only see imprecise, biased, and inferentially-derived representations of these inputs (7, 8). Additionally, deep neural networks exhibit features like adversarial perturbations highly uncharacteristic of human behavior (9), likely reflecting fundamental architectural differences between artificial and biological networks (10). If choosers optimize over both behavior and biological constraints, actual strategies may differ from the best-fitting product of biologically-unconstrained algorithms, potentially even falling outside the search space defined by psychological considerations alone.

Biological constraints also imply that individual choosers adapt to their unique context, as evident in two major remaining challenges. First is the impact of sequential effects in risky decision-making. Though theories have long asserted that choices are reference-dependent (11), recent research has begun to characterize how recent values (12, 13) and outcomes (14, 15) affect subsequent risk-taking from short (13, 15) to long timescales (12, 14). Second is the role of large individual differences in decision-making, addressed briefly in the supplemental material of Peterson et al. Analytical approaches on the scale of individuals must complement insights from larger-scale, aggregate analysis.

The powerful, data-driven tools leveraged by Peterson et al promise to significantly accelerate theoretical development in decision-making and beyond. However, this approach must ultimately engage directly with biological constraints and their consequences, and be used in conjunction with more granular analytical approaches for maximum scientific impact.

References

1. J. C. Peterson, D. D. Bourgin, M. Agrawal, D. Reichman, T. L. Griffiths, Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* 372, 1209-1214 (2021).
2. H. B. Barlow, in *Sensory Communication*, W. A. Rosenblith, Ed. (MIT Press, Cambridge, MA, 1961).
3. F. Attneave, Some informational aspects of visual perception. *Psychol Rev* 61, 183-193 (1954).
4. E. P. Simoncelli, B. A. Olshausen, Natural image statistics and neural representation. *Annu Rev Neurosci* 24, 1193-1216 (2001).
5. K. Louie, P. W. Glimcher, Efficient coding and the neural representation of value. *Ann N Y Acad Sci* 1251, 13-32 (2012).

6. M. M. Botvinick, Y. Niv, A. C. Barto, Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* 113, 262-280 (2009).
7. A. Rustichini, K. E. Conen, X. Cai, C. Padoa-Schioppa, Optimal coding and neuronal adaptation in economic decisions. *Nat Commun* 8, 1208 (2017).
8. R. Polania, M. Woodford, C. C. Ruff, Efficient coding of subjective value. *Nature neuroscience* 22, 134-142 (2019).
9. I. J. Goodfellow, J. Shlens, C. Szegedy, Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572, (2014).
10. F. Pulvermüller, R. Tomasello, M. R. Henningsen-Schomers, T. Wennekers, Biological constraints on neural network models of cognitive function. *Nature Reviews Neuroscience*, 1-15 (2021).
11. D. Kahneman, A. Tversky, Prospect theory - analysis of decision under risk. *Econometrica* 47, 263-291 (1979).
12. M. W. Khaw, P. W. Glimcher, K. Louie, Normalized value coding explains dynamic adaptation in the human valuation process. *Proc Natl Acad Sci U S A* 114, 12696-12701 (2017).
13. F. Rigoli, B. Chew, P. Dayan, R. J. Dolan, Learning Contextual Reward Expectations for Value Adaptation. *J Cogn Neurosci* 30, 50-69 (2018).
14. K. Juechems, J. Balaguer, M. Ruz, C. Summerfield, Ventromedial prefrontal cortex encodes a latent estimate of cumulative reward. *Neuron* 93, 705-714. e704 (2017).
15. H. R. Brooks, P. Sokol-Hessner, Quantifying the immediate computational effects of preceding outcomes on subsequent risky choices. *Scientific reports* 10, 1-10 (2020).