

Towards an atlas of canonical cognitive mechanisms

Angelo Pirrone¹ and Konstantinos Tsetsos^{2,3}

¹Centre for Philosophy of Natural and Social Science, London School of Economics, UK

²School of Psychological Science, University of Bristol, UK



³Department of Neurophysiology and Pathophysiology, University Medical Center
Hamburg-Eppendorf, Germany

Abstract

A central goal in Cognitive Science is understanding the mechanisms that underlie cognition. Here, we contend that Cognitive Science, despite intense multidisciplinary efforts, has furnished surprisingly few mechanistic insights. We attribute this slow mechanistic progress to the fact that cognitive scientists insist on performing underdetermined exercises, deriving overparametrised mechanistic theories of complex behaviours and seeking validation of these theories to the elusive notions of optimality and biological plausibility. We propose that mechanistic progress in Cognitive Science will accelerate once cognitive scientists start focusing on simpler explananda that will enable them to chart an atlas of elementary cognitive operations. Looking forward, the next challenge for Cognitive Science will be to understand how these elementary cognitive processes are pieced together to explain complex behaviour.

Keywords: optimality, biological plausibility, cognitive science, mechanisms, inference

Generating mechanistic theories of cognitive capacities is inevitably hard since theories in the empirical sciences are underdetermined by data. Researchers abduce theories not only on the basis of empirical observations but also on theoretical grounds, such as considerations about computability, parsimony, learnability, evolvability and more (Van Rooij, 2008; van Rooij & Baggio, 2021). We start by observing that in Cognitive Science, two notions – naturally linked to Marr’s computational and implementational levels of anal-

Angelo Pirrone  <https://orcid.org/0000-0001-5984-7853> Konstantinos Tsetsos  <https://orcid.org/0000-0003-2709-7634> This research was supported by a European Research Council Starting Grant under the European Union’s Horizon 2020 research and innovation program (Grant No. 802905, awarded to KT). No competing interests declared. Correspondence concerning this article should be addressed to Angelo Pirrone or Konstantinos Tsetsos. E-mails: a.pirrone@lse.ac.uk and k.tsetsos@bristol.ac.uk

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license (where permitted by UKRI, ‘Open Government Licence’ or ‘Creative Commons Attribution No-derivatives (CC BY-ND) license’ may be stated instead) to any Author Accepted Manuscript version arising.

ysis (Marr, 1982) – are routinely used to support the abduction of mechanistic theories: optimality and biological plausibility.

In this letter, we argue that seeking for ‘seals of approval’ from the auxiliary notions of optimality and biological plausibility is rather fruitless and revealing of a bigger issue. That is, theories in Cognitive Science are underdetermined by data to such a large extent, that auxiliary notions often become the primary criteria for theory selection (see related debates, Jones & Love, 2011; Lieder & Griffiths, 2020; Love, 2021; Rahnev & Denison, 2018). We postulate that a more productive trajectory should exactly address the issue of underdetermination drastically, by first focusing on simple behaviours that are amenable to mechanistic explanation; and then by carefully examining how these simple mechanisms can give rise to more complex behaviour.

Optimality as a seal of approval

Optimality is an elusive notion because its definition depends on a number of arbitrary assumptions. However, optimality can be used in a more constrained fashion by defining an optimal algorithm with regards to a certain criterion and assumptions. Typically, showing that an algorithm derived from or linked to optimality principles can quantitatively account for behavioural data, is perceived as a superior explanation in Cognitive Science, spanning more than one levels of Marr’s analyses, answering *what*, *why* and *how* questions concerning cognitive capacities. However, *a priori* assuming that an ‘optimal’ algorithm is the most satisfying computational account is unwarranted (Guest & Martin, 2021). Due to this misconception, the link between ‘optimal’ algorithms and the empirical reality often ends up being loose. That is, cognitive scientists readily validate ‘optimal’ algorithms once they quantitatively approximate a set of *generic* datapoints. However, the empirical tests that should probe the core assumptions of the algorithm are typically absent (van Rooij & Baggio, 2021).

For example, the drift-diffusion model (Ratcliff, 1978) is a decision-making algorithm that manages speed-accuracy trade-offs in a statistically optimal fashion, but does so narrowly only in specific and idealised environments (Moran, 2015). The narrow optimality that the drift-diffusion model exhibits, has wrongly served as ‘a seal of approval’. Cognitive and mathematical psychologists have focused on how well this model can explain countless *generic* datasets; the (often trivial) theoretical and empirical tests that could have falsified the model (or that could have ‘improved’ auxiliary aspects of the model without challenging core assumptions - see Lakatos, 1976) were postponed for almost four decades (Pirrone et al., 2018; Pirrone et al., 2022; Teodorescu et al., 2016). For instance, do decision-makers actually integrate *difference* in evidence? Do they even integrate evidence at all (Cisek et al., 2009; Edmunds et al., 2020)? Are there limits on the temporal window of information integration (Usher & McClelland, 2001)? Do people optimise speed-accuracy or speed-value trade-offs (Pirrone et al., 2014)? Interestingly, crucial observations obtained in ‘strong inference’ experiments (Platt, 1964) cast doubts on the validity of the algorithm.

Biological plausibility as a seal of approval

Biological plausibility is the extent to which a proposed mechanistic theory is consistent with the way the brain represents and processes information. It is reasonable to

assume that mechanistic theories must have viable neural implementations. However, akin to the notion of optimality, biological plausibility can be flexibly defined. A lot might be known about how single neurons work, but how computations scale-up in neural populations or how different brain regions communicate during complex tasks is a topic of intense investigation (Pessoa, 2022).

Even though it is a fuzzy concept, biological plausibility is often used as a seal of approval for mechanistic theories of complex behaviour (Busemeyer et al., 2019; Love, 2021; Roe et al., 2001; Usher & McClelland, 2004). For instance, neural network models often enjoy an elevated status as superior explanations of behaviour. However, the similarities between artificial and biological neural networks are rather superficial (see Gurney, 2018). The fact that artificial neural networks mimic instances of complex behaviour could just be a byproduct of their complexity and clever engineering techniques (Bowers et al., 2022). A common criticism is that neural network models end up as opaque as the brain itself.

The notion of biological plausibility can also mask the failure of certain mechanistic models in explaining and predicting behaviour. To illustrate, divisive normalisation is thought to be a canonical neural computation describing the way neurons represent inputs in the visual cortex (Louie et al., 2013). The idea that the inputs to a neuron are divisively normalised by their sum was recently applied in a decision-making model for 3-alternative value-based choices. In this setting, divisive normalisation predicts that the discriminability between the two strong alternatives is reduced when the weakest alternative increases in value. Nevertheless, the empirical robustness of this effect has been doubted (Gluth et al., 2020). More generally, the assumption that the neural computations underlying visual representation also underlie representation of value-based options is unwarranted.

Searching for canonical mechanisms

Reflecting on the examples above we postulate that if optimality or biological plausibility considerations weigh in heavily or even precede mechanistic descriptions, favouring a certain class of models is unavoidable and potentially misleading. Optimality and biological plausibility are flexible notions, and thus cannot be reliably used for theory selection, let alone as primary criteria. Ultimately, the strongest test of a mechanistic model lies in its ability to explain behaviour. Can't cognitive scientists focus directly on mechanistic explanations of behaviour? Adopting Marr's 'top-down' perspective, earlier influential frameworks (Anderson, 2013) discarded direct mechanistic explanations as being by-definition arbitrary. There are simply too many possible mechanistic explanations for a single behavioural phenomenon. We partly accept this assessment. Mechanistic explanations can be arbitrary, not by-definition, but when the targeted behaviours are overly complex. With the holy grail in Cognitive Science being the explanation of complex, sophisticated behaviours mechanistic stagnation is unsurprising.

We propose that before understanding the mechanisms via which humans do maths, play chess, or procrastinate, cognitive scientists need first to definitively understand the mechanisms via which humans retain, recall, forget, select and integrate information at the simplest level. Such mechanisms sound trivial but remain poorly understood. Mechanistic discovery of simple behaviours has been happening in a domain-specific fashion within Cognitive Psychology and Neuroscience but what is currently missing is theoretical unification by connecting findings across domains such as memory, attention, or simple choice in order

to chart the basic computational blocks of cognition. This endeavour should be intensified and assisted by advances in invasive and non-invasive neurophysiological techniques that permit the more detailed observation of information representation and processing in humans and other animals. While this approach is not immune to design choices and assumptions, highlighting what human cognition can do routinely across domains, mitigates the risks that conclusions are task-dependent or dependent upon specific assumptions. We believe that this approach lies somewhat closer to mechanisms in the trade-off between complexity and mechanistic observability, and as such is a promising plan to tackle fundamental questions pertaining cognition.

Our proposal is not merely a call for ‘keeping it simple’. The next challenge after understanding these simple mechanisms, is understanding how these simple mechanisms are pieced together during complex behaviours. Towards this direction, we advocate for a more sophisticated reductionist approach where the core aspects of complex tasks are abstracted away into less complex bespoke tasks that simultaneously probe more than one cognitive processes and their interactions (Brunton et al., 2013). Waskom et al. (2019) discuss examples from the literature, and propose practical recommendations regarding how experimental techniques from sensory research could be exported to fields such as the study of decision making and executive control using simple parametrised stimuli that allow to manipulate the timing and strength of information. Similarly, Tsetsos et al. (2012) built multi-attribute context effects into a seemingly innocuous evidence accumulation task by manipulating the temporal correlations among the incoming evidence samples. Obtaining ‘psychophysical analogues’ of behavioural phenomena that are typically described in consumer choice experiments facilitates a more direct mechanistic explanation of the factors that dictate our consumer choices.

Recent proposals, promising to accelerate our mechanistic understanding, involve both ‘computational-first’ (Lieder & Griffiths, 2020), and ‘implementational-first’ (Kriegeskorte & Douglas, 2018) approaches as ways of narrowing down the vast space of possible algorithms. We acknowledge that both these proposals are cautious in their usage of optimality and biological plausibility, beyond the superficial ‘seal of approval’ usage. However, fundamental theorising tools in these approaches (e.g., neural networks or Markov chain Monte Carlo sampling) assume that basic computational operations (e.g., argmax or integration), are implemented in stylised ways. We believe that these proposals should mitigate the tangible risk of these basic computational operations being implemented fundamentally differently in biological brains.

In conclusion, we contend that if both the processes and the inputs that feed into these processes are unknown, claims about cognitive capacities supported by optimality or biological plausibility claims are vacuous. We promote a bottom-up approach (Love, 2015; Waskom et al., 2019) more narrowly and intensely focused within Marr’s algorithmic level of analysis. Certainly, we do not claim that other levels of analysis should be dispensed with, but believe that discovering canonical algorithms can naturally feed back to the question of optimality (Moran & Tsetsos, 2018; Tsetsos et al., 2016) or neural implementation (Luyckx et al., 2020) in less ambiguous terms, by enabling the characterisation of the trade-offs that the cognitive system solves within its neural realisation.

References

- Anderson, J. R. (2013). *The adaptive character of thought*. Psychology Press.
- Bowers, J. S., Malhotra, G., Dujmović, M., Montero, M. L., Tsvetkov, C., Biscione, V., Puebla, G., Adolphi, F. G., Hummel, J., Heaton, R. F., et al. (2022). Deep problems with neural network models of human vision. *PsyArXiv*, *Published online November 14, 2022*. <https://psyarxiv.com/5zf4s/>.
- Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, *340*(6128), 95–98.
- Busemeyer, J. R., Gluth, S., Rieskamp, J., & Turner, B. M. (2019). Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends Cogn. Sci.*, *23*(3), 251–263.
- Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *J. Neurosci.*, *29*(37), 11560–11571.
- Edmunds, C. E., Bose, D., Camerer, C. F., Mullett, T. L., & Stewart, N. (2020). Accumulation is late and brief in preferential choice. *PsyArXiv*, *Published online July 29, 2020*. <https://psyarxiv.com/sa4zr>.
- Gluth, S., Kern, N., Kortmann, M., & Vitali, C. L. (2020). Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nat. Hum. Behav.*, *4*(6), 634–645.
- Guest, O., & Martin, A. E. (2021). On logical inference over brains, behaviour, and artificial neural networks. *Published online January 25, 2022*. <https://psyarxiv.com/tbmcg/>.
- Gurney, K. (2018). *An introduction to neural networks*. CRC press.
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. *Behav. Brain Sci.*, *34*(4), 169.
- Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nat. Neurosci.*, *21*(9), 1148–1160.
- Lakatos, I. (1976). Falsification and the methodology of scientific research programmes. In *Can theories be refuted?* (pp. 205–259). Springer.
- Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.*, *43*.
- Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proc. Natl. Acad. Sci. U.S.A.*, *110*(15), 6139–6144.
- Love, B. C. (2015). The algorithmic level is the bridge between computation and brain. *Top. Cogn. Sci.*, *7*(2), 230–242.
- Love, B. C. (2021). Levels of biological plausibility. *Philos. Trans. R. Soc. B*, *376*(1815), 20190632.
- Luyckx, F., Spitzer, B., Blangero, A., Tsetsos, K., & Summerfield, C. (2020). Selective integration during sequential sampling in posterior neural signals. *Cereb. Cortex*, *30*(8), 4454–4464.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman.

- Moran, R. (2015). Optimal decision making in heterogeneous and biased environments. *Psychon. Bull. Rev.*, *22*(1), 38–53.
- Moran, R., & Tsetsos, K. (2018). The standard bayesian model is normatively invalid for biological brains. *Behav. Brain Sci.*, *41*.
- Pessoa, L. (2022). The entangled brain. *J. Cogn. Neurosci.*, 1–12.
- Pirrone, A., Azab, H., Hayden, B. Y., Stafford, T., & Marshall, J. A. (2018). Evidence for the speed–value trade-off: Human and monkey decision making is magnitude sensitive. *Decision*, *5*(2), 129–142.
- Pirrone, A., Reina, A., Stafford, T., Marshall, J. A., & Gobet, F. (2022). Magnitude-sensitivity: Rethinking decision-making. *Trends Cogn. Sci.*, *26*(1), 66–80.
- Pirrone, A., Stafford, T., & Marshall, J. A. (2014). When natural selection should optimize speed-accuracy trade-offs. *Front. Neurosci.*, *8*, 73.
- Platt, J. R. (1964). Strong inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, *146*(3642), 347–353.
- Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behav. Brain Sci.*, *41*.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.*, *85*(2), 59–108.
- Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychol. Rev.*, *108*(2), 370.
- Teodorescu, A. R., Moran, R., & Usher, M. (2016). Absolutely relative or relatively absolute: Violations of value invariance in human decision making. *Psychon. Bull. Rev.*, *23*(1), 22–38.
- Tsetsos, K., Chater, N., & Usher, M. (2012). Saliency driven value integration explains decision biases and preference reversal. *Proc. Natl. Acad. Sci. U.S.A.*, *109*(24), 9659–9664.
- Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proc. Natl. Acad. Sci. U.S.A.*, *113*(11), 3102–3107.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychol. Rev.*, *108*(3), 550–592.
- Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychol. Rev.*, *111*(3), 757–769.
- Van Rooij, I. (2008). The tractable cognition thesis. *Cogn. Sci.*, *32*(6), 939–984.
- van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspect. Psychol. Sci.*, *16*(4), 682–697.
- Waskom, M. L., Okazawa, G., & Kiani, R. (2019). Designing and interpreting psychophysical investigations of cognition. *Neuron*, *104*(1), 100–112.