Commentary

Biological Psychiatry

Theory-Based Computational Psychiatry

Tiago V. Maia, Quentin J.M. Huys, and Michael J. Frank

Cannot we be content with experiment alone? No, that is impossible; that would be a complete misunderstanding of the true character of science…. Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house.

—Henri Poincaré, *Science and Hypothesis*, 1905 (Greenstreet WJ, transl, p. 157)

Philosophy [nature] is written in that great book which ever lies before our eyes—I mean the universe—but we cannot understand it if we do not first learn the language…in which it is written. This book is written in the mathematical language…without which one wanders in vain through a dark labyrinth.

—Galileo Galilei, *The Assayer*, 1623 (quoted in Burtt EA, *The Metaphysical Foundations of Modern Science*, p. 75)

Theory development is an intrinsic part of science. Radical empiricism is a logical impossibility: the number of phenomena that can be measured and manipulated is infinite, so the very selection of phenomena to investigate must be driven by a priori considerations. Loose facts, moreover, point to nothing but themselves; only theories, even if incipient, have explanatory and predictive power extending beyond prior observations. Yet, theory is sometimes seen with suspicion. Cajal, a giant in neuroscience history, wrote, "the theorist is a lazy person masquerading as a diligent one…a scholar's positive contribution is measured by the sum of the original data that he contributes…. Theories desert us, while data defend us" (1). Cajal wrote this text more than 2 centuries after the scientific revolution emphasized mathematical theories (consider Galileo's epitaph). Why? Later in the same paragraph, Cajal writes, "So many apparently conclusive theories…have collapsed in the last few decades! On the other hand, the well-established facts of anatomy and physiology…and *the laws and equations* of astronomy and physics remain" [(1), italics added]. Cajal was therefore not arguing against mathematical formulations of general principles—i.e., mathematical theories. Instead, he was arguing against vague verbal descriptions that constituted the theories in neuroscience at the time and that still characterize most theories in psychiatry, neuroscience, and related fields.

Theory-based computational psychiatry—the topic of this special issue of *Biological Psychiatry*—aims to use mathematically rigorous theories to help understand and hopefully better treat psychiatric disorders. Theories are unavoidable; theory-based computational psychiatry provides a framework to ensure that they are rigorous, consistent, and quantitative.

This special issue is witness to the excitement raised by computational psychiatry. Such excitement reflects the potential of computational psychiatry to meet three important challenges in psychiatry. First, psychiatric disorders involve complex interactions between phenomena at multiple levels of abstraction, ranging from the subcellular to the societal; computational techniques are uniquely suited to characterize such interactions (2–4). Second, psychiatry deals with remarkably complex phenomena; computational approaches enhance the comprehension, measurement, and prediction of such phenomena, including—critically for treatment development—prediction of the effects of manipulating variables (3). Third, the ever-increasing pace at which data are accumulated requires novel, more powerful computational tools. The first two aspects fall under the purview of theory-based computational psychiatry and are well illustrated in this issue. The third aspect falls under the purview of data- rather than theory-driven approaches, so it is not addressed in the issue, although it is also important.

The issue starts with two commentaries considering the way forward for computational psychiatry (5,6). In the first commentary, Moutoussis *et al.* (5) suggest approaches to promote the development of clinically useful applications of computational psychiatry, a topic that has recently received substantial attention (7–9). They suggest, among other directions, focusing on ecologically valid studies, relevant individual variability, and treatment processes. Their suggestion to use computational techniques to improve psychotherapy is particularly noteworthy: psychotherapy is basically a learning process, so it may benefit from the rich computational understanding of learning processes.

In the second commentary, Pine (6) focuses on the use of theory-driven, computationally defined mechanistic models to understand anxiety disorders. Using fear conditioning as an example, Pine addresses the usefulness of computational approaches to 1) facilitate—even force—precise thinking; 2) infer latent constructs; 3) solve the problem of task impurity (10); 4) disentangle multiple mechanisms that may produce the same effects [see, e.g., (11)]; and 5) guide experimental design to adjudicate among such mechanisms.

Following the commentaries, the issue contains four reviews (12–15)—three of which (13–15) propose novel theoretical perspectives—and two empirical reports (16,17). Together, these articles span a broad range of topics in psychiatry.

Voon *et al.* (12) review the literature on goal-directed (model-based) versus habitual (model-free) control and suggest that impaired model-based control may characterize compulsive behaviors cross-diagnostically. They support this argument by reviewing evidence for impaired model-based control in obsessive-compulsive disorder, alcohol

http://dx.doi.org/10.1016/j.biopsych.2017.07.016
Biological Psychiatry September 15, 2017; 82:382–384 www.sobp.org/journal
ISSN: 0006-3223

and stimulant dependence, and binge eating disorder, and by noting that a large factor-analytic study in the general population also found a selective relation to compulsivity (18). They also raise the crucial point that at least some of the tasks used to allegedly disentangle the model-based versus model-free systems—e.g., the two-step task (19)—engage the model-based system but may not be as sensitive to variations in the model-free system [although a recent variant of the two-step task addresses this limitation (20)]. What is classified as habit-based control in these tasks may therefore instead reflect superficial, partly incorrect, model-based inference. This possibility would explain the surprising findings, reviewed by Voon *et al.* (12), that increasing and decreasing dopamine in humans seems to make behavior more and less model-based, respectively. Rather than making behavior more model-based, increasing dopamine may simply make behavior that was already model-based more accurate by improving working memory (WM) and other executive functions, which would enhance the ability to make more complex inferences or better remember or use the model. Indeed, as shown by Collins *et al.* in this issue (16) and elsewhere (21), performance on even simple stimulus–response-like reinforcement learning tasks is strongly influenced by WM. Of course, substantial evidence shows that dopamine also affects model-free learning in humans, affecting learning from positive versus negative outcomes differentially (20,22).

In their prior work seeking to disentangle WM from model-free processes (21,23), Collins *et al.* used a task that, like the two-step task, may have been more sensitive to WM than to model-free processes. In their article in this issue (16), they present a task variant with similar sensitivity to WM but greater sensitivity to model-free processes. They found that, in healthy subjects, model-free learning was enhanced under high WM load. They also replicated their earlier finding in chronically medicated patients with schizophrenia (23) of profound deficits in WM contributions to learning but surprisingly spared model-free learning. Future work should investigate three possible explanations for this dissociation: 1) model-free learning truly is spared; 2) WM disturbances in patients mimic high-load conditions, thereby upregulating model-free learning and masking an inherent impairment; and 3) medications normalize model-free learning. Regardless, this novel task moves away from generalized deficits and presents an opportunity to study interacting cognitive and motivational systems in psychiatry.

Although, as reviewed by Voon *et al.* (12), most evidence links compulsivity to decreased model-based processes, not to increased model-free processes, substantial evidence implicates the model-free, habit-learning system in Tourette syndrome. Maia and Conceição (13) review this evidence, which suggests that tics are maladaptive motor habits. More importantly, they use current computational ideas about the specific roles of striatal phasic and tonic dopamine in action learning and invigoration, respectively, to suggest that increased striatal phasic and tonic dopamine in Tourette syndrome cause increased propensities to learn and express tics, respectively. They also show how the same computational ideas shed new light on the mechanisms of action of various medications used to treat Tourette syndrome.

Huys and Renz (14) focus on the problems that arise from cognitive-resource constraints. Model-based inference is too demanding computationally to be feasible in all but the simplest cases. Addressing this problem is the purview of meta-reasoning, which concerns the optimal allocation of cognitive resources: put simply, determining what one should think about to ensure one thinks of the best option. Unfortunately, meta-reasoning is even more intractable than model-based inference. Rather than allowing the problem to become compounded recursively, Huys and Renz (14) suggest that emotions may be used as approximate meta-reasoning strategies. They further argue that, together with a constructivist view of emotions as labels categorizing internal experiences, this perspective accounts for various aspects of emotion.

Petzschner *et al.* (15) propose a computational account of body control by the brain based on active inference [see also (24–27)]. Their framework unifies homeostasis and allostasis with probabilistic inference. In active inference, actions are aimed at reducing prediction errors (28). Their framework therefore suggests setting prior expectations to the desired physiological ranges (27); prediction errors then signal current or anticipated deviations from those values, which elicit homeostatic and allostatic control, respectively. Petzschner *et al.* (15) also consider the implications of these ideas for depression and autism spectrum disorders.

Huang *et al.* (17) report that subjects with high anxiety exhibit increased lose-shift behavior (switching after losses) even when it would be advantageous not to do so. This finding may point to a difficulty using statistical regularities to infer when to treat losses as spurious. However, these subjects' performance was not impaired, so they may instead have followed a different, but similarly adaptive, strategy.

Fully realizing the promise of theory-based computational psychiatry will be a long-term process. Progress will likely be gradual, rather than characterized by some watershed moment(s). Ensuring the long-term sustainability of this process without imperiling shorter-term advances will require a constant balancing act between developing theory-based approaches, seeking to apply them practically (8), and continuing to pursue "pragmatic" approaches, computational (7,29) or otherwise. Theory- and data-driven approaches should also be closely integrated (7). Regardless, the articles in this special issue demonstrate that progress is already here. Much work remains to be done, but one thing is certain: theory-based computational psychiatry is here to stay.

Acknowledgments and Disclosures

This work was supported by Swiss National Science Foundation Grant No. 320030L_153449/1 (to QJMH) and National Institute of Mental Health Grant No. R01 MH080066-01 (to MJF).

MJF is a consultant for F. Hoffmann-La Roche Pharmaceuticals. The other authors report no biomedical financial interests or potential conflicts of interest.

Article Information

From the Institute for Molecular Medicine (TVM), Faculty of Medicine, University of Lisbon, Lisbon, Portugal; Centre for Addictive Disorders (QJMH), Hospital of Psychiatry, University of Zurich, and Translational Neuromodeling Unit (QJMH), Institute of Biomedical Engineering, University of Zurich and the Swiss Federal Institute of Technology Zurich, Zurich,

Switzerland; the Department of Cognitive, Linguistic and Psychological Sciences (MJF), Department of Psychiatry and Human Behavior (MJF), and the Brown Institute for Brain Science (MJF), Brown University, Providence, Rhode Island.

TVM and QJMH contributed equally to this work.

Address correspondence to Tiago V. Maia, Ph.D., Institute for Molecular Medicine, Faculty of Medicine, University of Lisbon, Avenida Professor Egas Moniz, 1649-028 Lisbon, Portugal; E-mail: Tiago.V.Maia@gmail.com.

Received Jul 25, 2017; accepted Jul 25, 2017.

References

1. Ramón y Cajal S (1999): Advice for a Young Investigator [Swanson N, Swanson LW, trans]. Cambridge, MA: The MIT Press, 85–86 [original work published 1897; translation based on 4th edition, 1916].
2. Frank MJ (2015): Linking across levels of computation in model-based cognitive neuroscience. In: Forstmann BU, Wagenmakers E, editors. An Introduction to Model-Based Cognitive Neuroscience. New York: Springer, 159–177.
3. Maia TV (2015): Introduction to the series on computational psychiatry. Clin Psychol Sci 3:374–377.
4. Wang XJ, Krystal JH (2014): Computational psychiatry. Neuron 84:638–654.
5. Moutoussis M, Eldar E, Dolan RJ (2017): Building a new field of computational psychiatry. Biol Psychiatry 82:388–390.
6. Pine DS (2017): Clinical advances from a computational approach to anxiety. Biol Psychiatry 82:385–387.
7. Huys QJM, Maia TV, Frank MJ (2016): Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci 19:404–413.
8. Paulus MP, Huys QJM, Maia TV (2016): A roadmap for the development of applied computational psychiatry. Biol Psychiatry Cogn Neurosci Neuroimaging 1:386–392.
9. Huys QJM, Maia TV, Paulus MP (2016): Computational psychiatry: From mechanistic insights to the development of new treatments. Biol Psychiatry Cogn Neurosci Neuroimaging 1:382–385.
10. Wiecki TV, Poland J, Frank MJ (2015): Model-based cognitive neuroscience approaches to computational psychiatry: Clustering and classification. Clin Psychol Sci 3:378–399.
11. Maia TV, Cano-Colino M (2015): The role of serotonin in orbitofrontal function and obsessive-compulsive disorder. Clin Psychol Sci 3:460–482.
12. Voon V, Reiter A, Sebold M, Groman S (2017): Model-based control in dimensional psychiatry. Biol Psychiatry 82:391–400.
13. Maia TV, Conceição VA (2017): The roles of phasic and tonic dopamine in tic learning and expression. Biol Psychiatry 82:401–412.
14. Huys QJM, Renz D (2017): A formal valuation framework for emotions and their control. Biol Psychiatry 82:413–420.
15. Petzschner FH, Weber LAE, Gard T, Stephan KE (2017): Computational psychosomatics and computational psychiatry: Toward a joint framework for differential diagnosis. Biol Psychiatry 82: 421–430.
16. Collins AGE, Albrecht MA, Waltz JA, Gold JM, Frank MJ (2017): Interactions among working memory, reinforcement learning, and effort in value-based choice: A new paradigm and selective deficits in schizophrenia. Biol Psychiatry 82:431–439.
17. Huang H, Thompson W, Paulus MP (2017): Computational dysfunctions in anxiety: Failure to differentiate signal from noise. Biol Psychiatry 82:440–446.
18. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND (2016): Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. Elife 5:e11305.
19. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011): Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215.
20. Doll BB, Bath KG, Daw ND, Frank MJ (2016): Variability in dopamine genes dissociates model-based and model-free reinforcement learning. J Neurosci 36:1211–1222.
21. Collins AGE, Frank MJ (2012): How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci 35:1024–1035.
22. Maia TV, Frank MJ (2011): From reinforcement learning models to psychiatric and neurological disorders. Nat Neurosci 14:154–162.
23. Collins AGE, Brown JK, Gold JM, Waltz JA, Frank MJ (2014): Working memory contributions to reinforcement learning impairments in schizophrenia. J Neurosci 34:13747–13756.
24. Barrett LF, Simmons WK (2015): Interoceptive predictions in the brain. Nat Rev Neurosci 16:419–429.
25. Seth AK, Friston KJ (2016): Active interoceptive inference and the emotional brain. Phil Trans R Soc B 371:20160007.
26. Pezzulo G, Rigoli F, Friston K (2015): Active inference, homeostatic regulation and adaptive behavioural control. Prog Neurobiol 134: 17–35.
27. Stephan KE, Manjaly ZM, Mathys CD, Weber LAE, Paliwal S, Gard T, *et al.* (2016): Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. Front Hum Neurosci 10:550.
28. Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, O'Doherty J, Pezzulo G (2016): Active inference and learning. Neurosci Biobehav Rev 68:862–879.
29. Paulus MP, Huang C, Harlé KM (2016): Call for pragmatic computational psychiatry: Integrating computational approaches and risk-prediction models and disposing of causality. In: Redish AD, Gordon JA, editors. Computational Psychiatry: New Perspectives on Mental Illness (Strüngmann Forum Reports, vol. 20). Cambridge, MA: MIT Press, 259–274.