Trials for computational psychiatry

Quentin J.M. Huys[1] **and Michael Browning**[2,3]

1 Applied Computational Psychiatry Lab, Mental Health Neuroscience Department, Division of Psychiatry and Max Planck Centre for Computational Psychiatry and Ageing Research, Department of Imaging Neuroscience, Queen Square Institute of Neurology, University College London, UK **2** Department of Psychiatry, Oxford University, UK **3** Oxford Health NHS Trust, Oxford, UK

The authors contributed equally.

Accepted for publication in Nature Computational Science 28/8/2025.

Corresponding Author
Quentin JM Huys
Email: q.huys@ucl.ac.uk

Computational psychiatry is increasingly delivering causal evidence through a new focus on interventions research and clinical trials. This can improve patient outcomes through improved precision, repurposing, novel interventions, scaling of psychotherapy and improved translation.

Psychiatric disorders affect the higher functions of the brain. Motivation, inference, prediction, memory and perception all fall prey to psychiatric illnesses. Our understanding of these higher functions has been substantially advanced by computational methods over the past three decades. As such, it became natural to examine how computational methods might help understand and treat mental illnesses. Around a decade ago, interest in what became known as computational psychiatry started to rise[1]. This field has grown substantially, boasting its own journal, conferences, and courses. With the roots of computational psychiatry more in theoretical and neuroscience than clinical fields, research has often focused on the identification of symptomatic correlates of computational processes. As the field is moving towards clinical application, it is starting to engage with causation. Here, we briefly review the origins and outline how the field is moving towards causal approaches using clinical trials.

Origins: data- and theory-driven computational psychiatry

Computational psychiatry has often been subdivided into two broad branches[1]. Its theory-driven branch has translated the theoretical insights from neuroscience to questions relating to mental illness. The particular strength of theory-driven research is the strong link it provides between biology and higher functions of the brain, the most famous example probably being that of dopaminergic temporal prediction errors. Here, the theory allowed researchers to understand how cellular and even subcellular[2] processes implement a particular algorithm for solving an evolutionarily important problem: learning from delayed reinforcement. Examples where links between levels of description have been successfully identified now abound, ranging from vision to memory, beliefs, planning, concept formation and increasingly even emotions. Without the theory, there would be no such link between levels of description. The importance of such links cannot be overstated as—ultimately—any treatment targeted directly at the brain will require some link between the biological, implementation level, and the higher, cognitive levels at which patients experience the symptoms that bring them to the clinic.

Data-driven research in computational psychiatry acknowledges that the neurobiology of higher mental functions is unlikely to be simple enough to be fully understood. It instead concentrates on characterizing and exploiting complex patterns in rich datasets that are beyond the theories; but still address meaning-

ful problems. It has enabled researchers to characterise longitudinal patterns, associations and predictors to address clinical questions in a direct and immediate manner. Examples again abound, but include the prediction of treatment response from standard electroencephalography recordings and clinical measures[3].

While data- and theory-driven approaches differ in emphasis, they share close methodological and conceptual DNA. The methods used in theory-driven computational psychiatry are often derived from or substantially informed by machine learning. Conversely, machine-learning benefits from theoretical insights to develop useful feature extraction steps. Indeed, there is no such thing as theory-free data. After all, data-driven research into mental health focuses on data judged relevant to the question at hand, and this judgement essentially derives from theory.

The recent revolution in deep neural networks is driving major new advances. These data-driven tools have leapfrogged both theory-based models and standard data-driven approaches from machine-learning by extracting long-range, higher-order statistical structure from vast datasets. One notable application relevant to mental health is in the domain of language, which is central to the assessment of mental illness as the most direct observation of beliefs, schemas and the structure of thoughts. Large language models are enabling quantitative approaches linking language and concepts to neurobiology, e.g. in linking semantic processing to hippocampal (dys)function[4]. It is striking to reflect that a relatively atheoretical analysis should now underpin the most advanced analysis of content; where neuroscience had previously been broadly limited to the analysis of form.

Overall, however, both theory- and data-driven research in mental health has focused on correlational analyses, either via cross-sectional designs, or due to the focus on existing data. This has led to important insights, but improvement in patient outcomes will ultimately require the tackling of causal questions.

Paths towards causal impact

We suggest that the field should prepare to deliver causal evidence through a renewed focus on experimental medicine studies and clinical trials. While recent advances in causal inference from observational data offer some advantages over standard analytic approaches, none of these techniques provide the level of evidence on causality as randomization. Randomization reliably breaks the chain of confounding influences present in correlational studies. It establishes that an intervention has a consequence that is not due to some other confounder. In contrast to correlational evidence, such causal evidence has the potential to directly influence and improve clinical care and outcomes. We see several important causal paths for research in computational psychiatry to realize this. Most of these are amenable to causal experimental medicine studies, followed by clinical trials.

The first causal path is that of precision psychiatry[5]. Computational methods are ideally positioned to improve how existing interventions are deployed, e.g. by better identifying individuals likely to benefit from a given intervention. Such studies examine which computational mechanisms are engaged by specific interventions. There is a rich experimental medicine tradition to rely on here, and tasks typically used for computational psychiatry research are straightforward to adapt to these questions. Behavioural measures often take only a few minutes to complete, and can be deliverd online, enabling straightforward measurements of mechanisms in any clinical trial or longitudinal study. Where interventions are known to have specific effects[6], this opens up the possibility of identifying selective markers[7]. Once an intervention has shown to engage a particular mechanism reliably, this can be used for an enrichment or stratification trial[8], and in the future potentially for differential assignment studies.

The second path closely follows the first and concerns repurposing. Online tasks quantifying novel computational targets are cheap, scalable and accessible. As such, they can be added to longitudinal studies straightforwardly, either for specific promising agents, or potentially also more broadly to identify potential promising agents engaging specific computational mechanisms. This should open up the path towards experimental medicine studies and clinical trials for repurposed interventions.

The third path is to design novel interventions. Theory-driven computational psychiatry relies extensively on tasks to measures specific functions. Tasks can in turn be straightforwardly modified for interventions. Research in cognitive bias modification provides rich guidance here. Effects may be small[9] but nevertheless have a meaningful augmenting impact given their scalability[10]. Furthermore, cognitive bias modification research has shown that generalization from a simple task to real life is challenging, and motivates close integration with existing clinical approaches[11]. Computational psychiatry can provide tasks, and thereby interventions, which go beyond traditional cognitive bias modification targets and may be more mechanistically specific[12].

The fourth path arises at the intersection of psychotherapy and large language models[13]. These models help to adapt rigid scripts to specific individuals, thereby potentially augmenting the efficacy of the interventions, and helping to improve adherence. The study of therapist and patient language itself promises new insights into the active mechanisms of psychotherapies. We suggest that trials of these new interventions should be closely coupled to experimental methods to robustly and rapidly develop a better understanding of specific underlying computational mechanisms.

Finally it is important to remember that computational methods are comparatively amenable to translational research, enabling detailed neurobiological and preclinical investigations. The poster child of course is dopamine, where human and animal research go hand in hand[14]. Tasks probing computational processes are far more amenable to back-translation into animal models than clinical observations, and are inherently more valid.

BARRIERS

To succeed in these endeavours, we suggest that the field needs to address four barriers: measurement, outcomes, the gap between clinics and laboratories, and treatment studies.

First, measurement. Researchers are increasingly examining, characterising and optimizing measurement and psychometric properties of computational metrics. We are starting to routinely know what drives the reliability, identifiability and sensitivity of assessments, and how to improve each of these properties. Slowly, measurements are even starting to be normed at least for basic variables such as age and sex. An important property of computational measurements is the transparency of their statistical properties, which often derives from their being formulated as generative models. There has been much interest recently in this area, with substantial improvements in how, for instance, cognitive probes are designed[15], and in the use of open science practices.

Second, outcomes. Research in psychiatry is largely dependent on outcomes that were established half a century or more ago, often before the disorders as they are today were defined; and certainly well before the advances of computational research. Research has been hamstrung by these measures. The biological and computational irrelevance of existing outcome measures places stringent limits on how much progress can be made in identifying underlying computational or biological processes—akin to how the reliability of a measurement limits the causal and correlational effects it can reveal. New outcome measures need to be developed which are both relevant to health broadly, but also sensitive to underlying computational and biological processes. In fact, these dual requirements, combined with the scalability of both language and computational task assessments may provide a path forward: can large datasets be acquired which richly capture self-report and link this to computational mechanisms? Could such datasets enable us to carve illness at its joint? Regulatory bodies will rightly have a high threshold for accepting new outcome measures, as alterations in the outcomes has vast consequences for industry and hence for the entire health ecosystem. But we believe this change is coming.

Third, the people gap. Clinicians and basic researchers have different concerns and interests. Clinicians have rarely had the skills to properly develop computational research; and computational scientists have struggled to identify clinically meaningful questions. However, this is changing. The ubiquity of computational tools has enabled a new cohort of computationally highly skilled clinicians. Conversely,

the prominence of mental health research has led to a revolution in the understanding and awareness amongst computational researchers. Collaborations and shared training opportunities are becoming ever more common.

Finally, treatment studies. Performing research on interventions short of trials needs to be very substantially facilitated. Regulatory clinical trials are expensive, demanding and slow, and the imposition of these demands on the developmental pipeline is profoundly stifling. To ensure that regulatory trials fail less frequently, computational measures need to be robustly validated in advance. This means that a testbed is required which enables fast, scalable validation in real, clinical, longitudinal and interventional settings. Cognitive neuroscience has been greatly aided by the advent of online testing, apps and online recruitment. Expanding such digital means to facilitate collection of data from patients undergoing treatments at scale could revolutionize the field, and any such opportunities need to be pursued with great persistence.

OUTLOOK

Much work remains to be done before trials in computational psychiatry can start to shape clinical practice. However, a collaborative focus causal research leading to trials, we believe, should greatly accelerate the development of improved treatments and hopefully strengthen societal support for science.

ACKNOWLEDGEMENTS

QJMH has received research grant funding from NIHR, Carigest S.A., German Research Foundation, Koa Health, Swiss National Science Foundation and Wellcome Trust (221826/Z/20/Z and 226790/Z/22/Z). QJMH acknowledges support by the NIHR University College London Biomedical Research Centre (NIHR UCLH BRC). MB was supported by the Office for Life Sciences and the National Institute for Health and Care Research (NIHR) Mental Health Translational Research Collaboration, hosted by the NIHR Oxford Health Biomedical Research Centre and by the NIHR Oxford Health Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health and Social Care.

QJMH has obtained fees and options for consultancies for Aya Technologies and Alto Neuroscience. MB has received consulting fees from J&J, Engrail Therapeutics, Alto Neuroscience, Empyrean Neuroscience, Boheringer Ingelhiem, CHDR, and travel expenses from Lundbeck. He was previously employed by P1vital Ltd.

CONTRIBUTIONS

The authors contributed equally.

REFERENCES

[1] Huys, Q.J.M., Maia, T.V. & Frank, M.J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci* **19**, 404–413 (2016).

[2] Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J. *et al.* Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* **525**, 243–246 (2015).

[3] Chekroud, A.M., Hawrilenko, M., Loho, H., Bondar, J., Gueorguieva, R. *et al.* Illusory generalizability of clinical prediction models. *Science (New York, N.Y.)* **383**, 164–167 (2024).

[4] Nour, M.M., McNamee, D.C., Liu, Y. & Dolan, R.J. Trajectories through semantic spaces in

schizophrenia and the relationship to ripple bursts. *Proceedings of the National Academy of Sciences of the United States of America* **120**, e2305290120 (2023).

[5] Williams, L.M. & Hack, L.M., eds. *Precision Psychiatry: using neuroscience insights to inform personally tailored, measurement-based care.* American Psychiatric Association Publishing, Washington D. C., USA (2022).

[6] Furukawa, T.A., Tajika, A., Toyomoto, R., Sakata, M., Luo, Y. *et al.* Cognitive behavioral therapy skills via a smartphone app for subthreshold depression among adults in the community: the resilient randomized controlled trial. *Nature medicine* **31**, 1830–1839 (2025).

[7] Norbury, A., Hauser, T., Fleming, S., Dolan, R. & Huys, Q.J.M. Different components of cognitive-behavioural therapy affect specific cognitive mechanisms. *Science Advances* **10**, eadk3222 (2024).

[8] Browning, M., Bilderbeck, A.C., Dias, R., Dourish, C.T., Kingslake, J. *et al.* The clinical effectiveness of using a predictive algorithm to guide antidepressant treatment in primary care (predict): an open-label, randomised controlled trial. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **46**, 1307–1314 (2021).

[9] Martinelli, A., Grüll, J. & Baum, C. Attention and interpretation cognitive bias change: A systematic review and meta-analysis of bias modification paradigms. *Behaviour Research and Therapy* **157**, 104180 (2022).

[10] Rinck, M., Wiers, R.W., Becker, E.S. & Lindenmeyer, J. Relapse prevention in abstinent alcoholics by cognitive bias modification: Clinical effects of combining approach bias modification and attention bias modification. *Journal of Consulting and Clinical Psychology* **86**, 1005–1016 (2018).

[11] Ashar, Y.K., Gordon, A., Schubiner, H., Uipi, C., Knight, K. *et al.* Effect of pain reprocessing therapy vs placebo and usual care for patients with chronic back pain: A randomized clinical trial. *JAMA psychiatry* **79**, 13–23 (2022).

[12] Norbury, A., Dercon, Q., Hauser, T.U., Dolan, R.J. & Huys, Q.J.M. Learning training as a cognitive restructuring intervention. *Biological psychiatry: Cognitive neuroscience and neuroimaging*. page Online ahead of print (2025).

[13] He, Y., Yang, L., Qian, C., Li, T., Su, Z. *et al.* Conversational agent interventions for mental health problems: Systematic review and meta-analysis of randomized controlled trials. *Journal of medical Internet research* **25**, e43862 (2023).

[14] Schmack, K., Bosc, M., Ott, T., Sturgill, J.F. & Kepecs, A. Striatal dopamine mediates hallucination-like perception in mice. *Science (New York, N.Y.)* **372** (2021).

[15] Huys, Q. & Paulus, M. Special issue on reliable mechanisms for translational applications. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **8**, 778–779 (2023).