14TLP0155_Fletcher

Computational psychiatry: a Rosetta Stone linking the brain to mental illness

Philip R Corlett, Paul C Fletcher,*

Yale University, Department of Psychiatry, Ribicoff Research Facility. New Haven, CT, USA (PR Corlett);
University of Cambridge, Department of Psychiatry, Addenbrooke's Hospital, Cambridge; Cambridge and Peterborough Foundation Trust; Wellcome-MRC Behavioural and Clinical Neuroscience Institute. (PC Fletcher)

Correspondence to: Paul C Fletcher, University of Cambridge, Herchel Smith Building, Forvie Site, Robinson Way, Cambridge, CB2 0SZ UK

pcf22@cam.ac.uk

Although psychiatry has a rich variety of models, most fail to span biological, psychological, and social domains. Computational psychiatry (panel) offers simple, direct ways of uniting these levels of explanation and analysis by providing notions that are applicable to each one.[1] Computational psychiatry holds, we argue, serious promise for development of comprehensive explanations and treatments for psychiatric illness; it can provide an account of how a person interacts with and is influenced by the world and so should be closely aligned to cognitive approaches in the investigation and treatment of mental illness.[2] We suggest much could be gained from harnessing the remarkable developments in computational neuroscience and the growing success in application of these techniques to our understanding of brain processes, cognition, and social interaction, and applying them to mental illness. This is the central goal of computational psychiatry, and the aim of this brief Viewpoint is to show its potential value. We give three linked examples relating to cognitive modelling of psychosis, neurochemical investigations, and integration of pharmacological and cognitive interventions. Although our focus is on psychosis and uses perspectives emerging from the field of learning and reinforcement, the points raised are applicable to other mental illnesses, and a great deal has been achieved with economic and game-theoretic perspectives.[3]

Cognitive neuropsychiatry seeks an understanding of mental symptoms in terms of healthy psychological functions (eg, exploration of auditory hallucinations in terms of processes related to monitoring of inner speech).[2] Computational psychiatry provides a means of refining and testing such theories, to identify key sub-processes and, in doing so, to characterise deviations more specifically. For example, the computational approach characterises learning through several parameters, such as motivation, value representation, learning rate, confidence, exploratory behaviour, prediction error, and meta-learning.[1] Armed with the ability to identify and measure these parameters, computational psychiatrists are in a good position to get to the heart of abnormal task performance and to link cognitive changes to behaviour and symptoms.

As an example, the jumping to conclusions (JTC) bias is invoked as an explanatory mechanism in models of delusions.[4] When people with schizophrenia (who have or are prone to delusions) make decisions in an uncertain setting, they tend to draw on less information in reaching that decision than does a person without schizophrenia. They might also ignore pieces of evidence that contradict that decision. Langdon and colleagues suggested that this "bias towards hasty decisions may contribute to the formation of delusions…"[5]. However, although the bias is invoked in this way to explain delusions, it is not specific[6] and, alone, its explanatory value is very restricted. Delusions are beliefs reached without strong objective evidence supporting them. Thus, portrayal of people with delusions as having a JTC bias is a redescription rather than an explanation. We should explore the underlying information processing in terms that relate to other levels of analysis. Computational psychiatry is thus essential if progress is to be made.

By exploring the bias more deeply, in terms of underlying information processing, psychiatrists can move towards other levels of enquiry and to the brain. For example, the JTC bias might arise because of altered noise in decision making.[7] Ideas of noise in a system that is concerned with prediction and inference link directly to predictive coding theory, which asserts that the brain strives to predict the world and optimise its predictions by minimising errors.[8] Within this framework, the brain is a model of its world, recapitulating in its structure and function past experiences of regularities. The brain infers the causes of its inputs on the basis of such experience and tests its inferences by making predictions about ensuing inputs. To function successfully in this way, actual inputs should be compared with expectations. Mismatches are signalled and minimised in several possible ways: they can be ignored (experiencing what we expect rather than what is presented by our senses), changed (altering the world to fit expectation), or updated (a changed model of the world). Error minimisation can be seen as a core principle of survival (avoid surprises to stay alive)[8], and we propose that anomalies in the genesis

of prediction error, its nature, and our responses to it can explain the emergence of psychosis. In the early phases of psychosis, the world becomes a strange, unpredictable place, with complex, distressing, and socially isolating experiences.[9] Predictive coding models explain these experiences as attempts to account for and minimise uncertainty.[10]

The idea central to predictive coding (that present input is shaped and interpreted by appealing to stored experience) is not new to psychiatry but computational psychiatry offers richer perspectives on phenomena such as the JTC bias and, crucially, links them (figure 1) to underlying brain processes without ignoring or sacrificing high level factors such as emotions, interpersonal interactions, and culture.[12] This information should be especially useful to the clinician, who struggles with the challenge of understanding and treating symptoms by combination of knowledge of the neurochemical, cognitive, and social domains. Computational Psychiatry is not just an academic exercise: if, for example, the JTC bias is a target for cognitive therapeutic intervention, a profound difference would exist between an approach guided by the knowledge that the anomaly occurred at the point of decision making and one that presumed an antipathy towards gathering of information.

The previous example offers a perspective on psychosis that is predominantly related to cognition and information processing. We can also, albeit imprecisely, re-express the predictive coding model in pharmacological terms, relating perturbed dopamine transmission to psychosis via disrupted learning mechanisms,[12–14] thus linking psychotic experiences to neurochemistry.[10] In simplified terms, predictive coding includes top-down signalling via N-methyl-D-aspartate (NMDA) glutamate receptors and bottom-up (prediction error) signalling via α-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid (AMPA) receptors.[15] Dopamine modulates the prediction error signal, enhancing its gain.[16] We speculate that the system's fundamental role of minimisation of prediction error and uncertainty could be perturbed by altered NMDA-mediated processing (leading to a failure of previous beliefs to constrain present experience) and altered AMPA processing associated with a change in bottom-up messages (leading to noise in the system that previous beliefs would have difficulty in minimising).[10] Moreover, altered dopamine signalling could give undue weight to these bottom-up signals. Pharmacological modelling of delusions with the psychotomimetic NMDA receptor antagonist ketamine provides a clue to how those perturbations are manifest biologically, through a loss of of the influence top-down priors (NMDA receptor blockade) and an enhancement of prediction error signalling (increased AMPA receptor transmission).[17]

With a growing and ever more nuanced understanding of the association between pharmacology and cognition, interventions could be designed empirically and particular patients could be offered particular treatment approaches dependent on the points at which they diverge from the model. For example, we know that key drivers of earning - prediction error and attentional salience[18] are mediated by separate but interacting neurotransmitter systems (dopamine and acetylcholine)[20,21]. Disruptions to both processes (prediction error and attentional salience attribution) can manifest in behaviour as aberrant learning and, ultimately, delusional beliefs.[10] A patient might experience delusions because of a disruption to either system.[10] The behavioural, neural, and algorithmic metrics derived from fitting a learning model to that person's brain and behavioural data would allow a clinician to ascribe their problem to either prediction error or attentional salience. A drug (or psychological treatment or combination treatment) could then be chosen for that patient that addressed their specific neurochemical dysfunction. In this way, Computational psychiatry allows us to marshal the advances in neuroscience and computation to tackle the heterogeneity of underlying pathophysiology that attends serious mental illnesses.

One puzzling characteristic of delusions is that, although they are fixed and impervious to contradictory evidence, they can nevertheless show remarkable elasticity, expanding to incorporate contradictory evidence in such a way that the delusional belief is strengthened rather than relinquished. This characteristic can be considered within the predictive coding framework, where prediction error might lead to an updated belief or might be suppressed.[22] We can relate this to animal conditioning: when an animal is exposed to pairings between an environmental stimulus (eg, a tone) and a salient event (eg, an electric shock), it learns that the tone predicts the shock. If the tone is then presented in the absence of shock, extinction learning occurs. The animal's behaviour can be thought of as being mediated by a competition between its previous learning and its new experiences.[23] This balance between extinction and reconsolidation of memories might be disrupted in people with psychosis, such that there is reduced extinction learning and possibly even belief-strengthening on extinction trials.[22]

This theory has important therapeutic potential. Cognitive behavioural therapy could guide new extinction learning.[22] The therapist encourages the patient to consider and adopt alternative explanations for their psychotic experiences. Viewed according to the computational tenets outlined in figure 1, cognitive behavioural therapy would be acting at the algorithmic level to encourage a different balance between representations, such that a new, non-psychotic belief prevails. In a 2011 investigation[24], participants given D-cycloserine (which boosts NMDA receptor function) after cognitive behavioural therapy showed an enhanced adoption of new explanatory beliefs compared with those who received placebo. This finding is readily comprehensible within a computational framework but less easily so within more restricted single level models.

Computational psychiatry has limitations. A recent review[1], stated that "much of the literature that substantiates the points we make has yet to appear". Although this fact is undeniable, we suggest that there are already profound insights enshrined in the computational psychiatry approach that are of direct relevance to the practising psychiatrist. These take the form of a desire to harness the emerging insights from cognitive neuroscience, a belief that such insights are of far more than academic interest and a dissatisfaction with explanations that fail to go beyond single levels (whether neurobiological, cognitive, or social).. Computational psychiatry is about modelling the brain in the world (indeed, it is about modelling how the brain models the world or even about modelling how the brain models how other brains model the world). As George Box said, all models are wrong, the question is how useful they are.[25] So, Bayesian analysis should not be conflated with Bayesian processing.[26] That is, Bayesian methods can be used to analyse data without the neural or cognitive system that engender those data necessarily being Bayesian.[26] And even if a Bayesian model provides a good fit for the data generated, that does not mean that the system functions in a Bayesian manner. There have been many examples in which Bayesian models can account for the neural data.[27] However, such examples are subtly different from proof of how the system works. One prediction of the Bayesian model of psychosis that was borne out by the data is the division of labour in glutamate receptor signalling between bottom-up (AMPA) and top-down (NMDA). This prediction was supported with neurophysiological recording in awake, behaving monkeys.[28] Clearly, more studies need to be done before this scheme determines clinical practice in psychiatry. However, computational psychiatry provides a means through which hypotheses can be generated and tested and, as such, it provides a hypothetico-deductive roadmap toward its own clinical implementation.[29]

A computational approach to psychiatry might seem something of an indulgence, arcane, abstract, and remote from the questions and challenges that face clinicians and patients. Moreover, it carries unfortunate connotations of the mechanical, the disembodied, and the emotionless. Such a view is wrong. In more than most fields, Computational psychiatry strives towards a truly biopsychosocial perspective by showing how each level of analysis (through neurons, circuits, cognitive processes, social interactions) can only be fully understood by characterising its association to other levels. Thus, the predictive coding model we have described and in figure 1 has, at its core, the idea that the brain tries to make inferences about its world and become a model of its

world. The brain is embedded in an environment that should be described in terms that include the social and the environmental. The brain makes the world and the world makes the brain.[30] A view of mental illness that fails to take this balance into account will be incomplete. Computational psychiatry recognises this danger and strives to avoid it.

Panel

Computational psychiatry in a nutshell

At its heart, computational psychiatry (computational psychiatry) holds that human experience, decision making, and behaviour, in all their complexity, might be understood in terms of how we build mental representations of the states of the world and act to influence those states as best as possible. Psychiatric illness and distress might be considered in terms of a failure to achieve this optimum interaction, and the challenge faced by computational psychiatry is to identify and quantify this suboptimal state. What is optimum might be established by the states and values of the individual and the state of the world. Because the world includes other minds and mental representations as well as complex social structures, computational psychiatry strives for the richest perspective possible. Computational psychiatry cannot ignore any level of analysis because each level contributes to the problems faced (and the solutions proposed) by the brain.

The approach recognises that even small detrimental changes in information processing can be devastating but that, nonetheless, many junctures exist at which intervention might be possible; we can change the environment for people who are in mental distress (perhaps assisting them to find housing or employment), or we can try to enable them to represent appropriate parameters correctly and use those representations to guide their behaviour optimally (eg, with cognitive behavioural therapy). Interventions such as these are already integral to psychiatry, but the key idea is that, by understanding the mapping from the problem being solved to the machinery of problem solving, we can more systematically characterise the problems and intervene at all levels to ameliorate disruptions to information processing and the effects of these disruptions.

Computational psychiatry and its debt to cognitive neuropsychiatry

Computational psychiatry has its antecedents in cognitive neuropsychiatry.[2] A review of the field defined computational psychiatry (in part) as "the use of formal models of brain function to characterise the mechanisms of psychopathology,"[31] which overlaps very clearly with the approach proposed by Halligan and David: "let us take the best cognitive models for a range of normal psychological functions and treat them as if psychiatric phenomena fall within their ambit"[2]. Computational psychiatry, like cognitive neuropsychiatry, involves specification of mental symptoms as departures from healthy processing. However, computational psychiatry includes a formal specification of the processes and the parametric means through which they deviate from healthy processing; in short, unlike cognitive neuropsychiatry, computational psychiatry is both qualitative (which parameters change?) and quantitative (by how much?).[1]

Figure legends

***Figure*: Levels of explanation**

The terms computational level, algorithmic level, and physical level are taken from an influential account of computing by David Marr and Tomaso Poggio.[11]

The physical anatomy of the brain embodies Bayesian mechanisms of top-down prediction of sensory inputs and bottom-up prediction errors. This finding allows us to solve the ill-posed problems of perceptual inference and decision-making in noisy environments such that we can respond adaptively and flexibly to the contingencies in our world.

H=hypothesis. D=data. P(H)=prior probability of H – the probability of H before seeing D. P(D)=probability of observing data, D. P(D|H)=likelihood – the probability of seeing D given H is true. P(H|D)=the posterior probability – the probability of hypothesis given the data. NMDA= N-methyl-D-aspartate. AMPA=α-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid.

1 Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci* 2012; **16:** 72–80.
2 Halligan PW, David AS. Cognitive neuropsychiatry: towards a scientific psychopathology. *Nat Rev Neurosci* 2001; **2:** 209¯15.
3 Kishida KT, King-Casas B, Montague PR. Neuroeconomic approaches to mental disorders. *Neuron* 2010; **67:** 543¯54.
4 Garety PA, Kuipers E, Fowler D, Freeman D, Bebbington PE. A cognitive model of the positive symptoms of psychosis. *Psychol med* 2001; **31:** 189¯95.
5. Langdon R, Still M, Connors MH, Ward PB, Catts SV. Jumping to delusions in early psychosis. *Cogn Neuropsychiatry* 2014; **19:** 241¯56.
6 Wittorf A, Giel KE, Hautzinger M, et al. Specificity of jumping to conclusions and attributional biases: a comparison between patients with schizophrenia, depression, and anorexia nervosa. *Cogn Neuropsychiatry* 2012; **17:** 262¯86.
7 Moutoussis M, Bentall RP, El-Deredy W, Dayan P. Bayesian modelling of Jumping-to-Conclusions bias in delusional patients. *Cogn Neuropsychiatry* 2011; **16:** 422¯47.
8 Friston K. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci* 2009; **13:** 293¯301.
9 Gross G, Huber G. Sensory disorders in schizophrenia. *Arch Psychiatr Nervenkr* 1972; **216:** 119¯30.
10 Corlett PR, Taylor JR, Wang XJ, Fletcher PC, Krystal JH. Toward a neurobiology of delusions. *Prog Neurobiol* 2010.
11 Marr D, Poggio, T. From understanding computation to understanding neural circuitry. *Neurosciences Res Prog Bull* 1977; **204**: 301¯28.
12 Kapur S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry* 2003; **160:** 13¯23.
13 Miller R. Schizophrenic psychology, associative learning and the role of forebrain dopamine. *Med Hypotheses* 1976; **2:** 203¯11.
14 Gray JA, Feldon J, Rawlins JNP, Hemsley D, Smith, A.D. The neuropsychology of schizophrenia. *Behav Brain Sci* 1991; **14**: 1¯84.
15 Friston K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 2005; **360:** 815¯36.
16 Friston KJ, Shiner T, FitzGerald T, et al. Dopamine, affordance and active inference. *PLoS Comput Biol* 2012; **8:** e1002327.
17 Corlett PR, Frith CD, Fletcher PC. From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology (Berl)* 2009; **206:** 515¯30.
18 Schultz W, Dickinson A. Neuronal coding of prediction errors. *Annu Rev Neurosci* 2000; **23**: 473¯500.
19 Dickinson A. The 28th Bartlett Memorial Lecture. Causal learning: an associative analysis. *Q J Exp Psychol B* 2001; **54:** 3¯25.
20 Chiba AA, Bucci DJ, Holland PC, Gallagher M. Basal forebrain cholinergic lesions disrupt increments but not decrements in conditioned stimulus processing. *J Neurosci* 1995; **15:** 7315¯22.
21 Bao S, Chan VT, Merzenich MM. Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* 2001; **412:** 79¯83.
22 Corlett PR, Krystal JH, Taylor JR, Fletcher PC. Why do delusions persist? *Front Hum Neurosci* 2009; **3:** 12.
23 Eisenhardt D, Menzel R. Extinction learning, reconsolidation and the internal reinforcement hypothesis. *Neurobiol Learn Mem* 2007; **87:** 167¯73.
24 Gottlieb JD, Cather C, Shanahan M, Creedon T, Macklin EA, Goff DC. D-cycloserine facilitation of cognitive behavioral therapy for delusions in schizophrenia. *Schizophr res*. 2011; **131:** 69¯74.
25 Box GE, Draper, N. Empirical Model-Building and Response Surfaces. New York: John Wiley & Sons, 1987.
26 Bowers JS, Davis CJ. Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin* 2012; **138:** 389¯414.
27 Deneve S. Bayesian spiking neurons II: learning. *Neural Comput* 2008; **20:** 118¯45.

28 Self MW, Kooijmans RN, Super H, Lamme VA, Roelfsema PR. Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proceedings of the National Academy of Sciences of the United States of America* 2012; **109:** 11031–36.
29 Stephan KE, Mathys C. Computational approaches to psychiatry. *Current opinion in neurobiology* 2014; **25**: 85–92.
30 Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and brain sciences* 2013; **36:** 181–204.
31 Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry* 2014; **1:** 148–158.