OPEN ACCESS

REVIEW

Computational Psychiatry: towards a mathematically informed understanding of mental illness

Rick A Adams,[1,2] Quentin J M Huys,[3,4] Jonathan P Roiser[1]

[1]Institute of Cognitive Neuroscience, University College London, London, UK
[2]Division of Psychiatry, University College London, London, UK
[3]Translational Neuromodeling Unit, University of Zürich and Swiss Federal Institute of Technology, Zürich, Zürich, Switzerland
[4]Department of Psychiatry, Psychotherapy and Psychosomatics, Hospital of Psychiatry, University of Zürich, Zürich, Switzerland

Correspondence to
Dr Rick A Adams, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3BG, UK;
rick.adams@ucl.ac.uk

Received 2 March 2015
Revised 29 May 2015
Accepted 19 June 2015
Published Online First 8 July 2015

ABSTRACT

Computational Psychiatry aims to describe the relationship between the brain's neurobiology, its environment and mental symptoms in computational terms. In so doing, it may improve psychiatric classification and the diagnosis and treatment of mental illness. It can unite many levels of description in a mechanistic and rigorous fashion, while avoiding biological reductionism and artificial categorisation. We describe how computational models of cognition can infer the current state of the environment and weigh up future actions, and how these models provide new perspectives on two example disorders, depression and schizophrenia. Reinforcement learning describes how the brain can choose and value courses of actions according to their long-term future value. Some depressive symptoms may result from aberrant valuations, which could arise from prior beliefs about the loss of agency ('helplessness'), or from an inability to inhibit the mental exploration of aversive events. Predictive coding explains how the brain might perform Bayesian inference about the state of its environment by combining sensory data with prior beliefs, each weighted according to their certainty (or precision). Several cortical abnormalities in schizophrenia might reduce precision at higher levels of the inferential hierarchy, biasing inference towards sensory data and away from prior beliefs. We discuss whether striatal hyperdopaminergia might have an adaptive function in this context, and also how reinforcement learning and incentive salience models may shed light on the disorder. Finally, we review some of Computational Psychiatry's applications to neurological disorders, such as Parkinson's disease, and some pitfalls to avoid when applying its methods.

To cite: Adams RA, Huys QJM, Roiser JP. *J Neurol Neurosurg Psychiatry* 2016;**87**:53–63.

INTRODUCTION

Computational Psychiatry aims first to model the computations that the brain performs—that is, the brain's solutions to the problems it faces—and second to thereby understand how the 'abnormal' perceptions, thoughts and behaviours that are currently used to define psychiatric disorders relate to normal function and neural processes. By formalising mathematically the relationship between symptoms, environments and neurobiology, it hopes to provide tools to identify the causes of particular symptoms in individual patients.

Computational Psychiatry is at least partially motivated by the shortcomings of the current psychiatric classification systems (the Diagnostic and Statistical Manual of Mental Disorders, or DSM-5,[1] and the International Classification of Diseases, or ICD-10[2]), in which the symptoms entail the diagnosis and which lack mechanistic explanations for mental symptoms. The reliability of diagnostic systems was 'bought at the price of validity'[3]—meaning clinicians have some confidence that given a set of symptoms they would all make a consistent diagnosis, but no confidence that that diagnosis corresponds to a single biological or psychological entity, or that it can predict the outcome of either the illness or a given treatment. Likewise, the biopsychosocial model of mental illness[4] has had great success in helping clinicians understand illness at a human level, but as a causal account it fails: its constituent parts (particularly the biological and psychosocial) are separated by a wide explanatory gap.

The National Institute of Mental Health (NIMH) generated the Research Domain Criteria (RDoC[5]) in an attempt to revive psychiatric classification with a bracing dose of mechanistic validity. The RDoC consists of five 'domains' of mental functions that are each described at multiple levels or 'units of analysis': the hope is that these units will yield biomarkers to distinguish normal and abnormal functioning. In principle, this approach has several advantages, but we note that the current RDoC (it is a working document) views psychiatric disorder—including its social risk factors—through a very biological lens. Indeed, its units of analysis step from 'genes' to 'molecules' to 'cells' to 'circuits' to 'physiology', and then leap straight to 'behaviour'. Computational Psychiatry provides some of the tools to link these levels.

Numerous authoritative reviews of initial developments in Computational Psychiatry already exist,[6–15] alongside pioneering work by Hoffman,[16] Cohen[17] and many others. In this article, we look towards the future and—using examples from depression and schizophrenia—illustrate Computational Psychiatry's potential for reconceptualising psychiatric disorders and generating new hypotheses. Prior to this, we briefly rehearse the advantages in adopting a Computational Psychiatry approach.

Computational Psychiatry unites many levels of description

Computational Psychiatry's organising principles arose in computational neuroscience, when Marr[18] identified three levels at which the problems solved by the brain may be described. At a 'computational' level, the formal nature of the problem has to be described: What are the mathematical and statistical issues involved? What solutions do these issues allow? The 'algorithmic' level describes the method of solving the problem. This may be an

approximation or a much more complex procedure. The 'implementational' level describes the physical realisation of this method: How does coordinated activity in neurons or brain circuits encode these algorithms?

Critically, these three levels are not entirely independent. Although any algorithm could be implemented physiologically in many ways, constraints at one level have implications at other levels: Some computations (eg, high-dimensional integrals) may be very laborious for neural systems, so algorithmic approximations become necessary. System failures caused by complex computational problems can therefore provide important clues about underlying algorithms.

In the biological sphere, this simple trinity is crossed with other relevant levels of description. For example, the implementational level itself can be decomposed into RDoC's various 'units of analysis', from genes to physiology as well as (we would argue) social interaction.[8 19] With respect to most mental disorders, Computational Psychiatry lies at the nexus of these descriptive levels and makes them explicit.

Computational Psychiatry is mechanistic and rigorous

Computational Psychiatry is mechanistic in a way that the DSM-5, ICD-10 and biopsychosocial model can never be, thanks to its use of generative models. A generative model is a probabilistic description of how high-level causes actually generate low-level data (in contrast, a discriminative model merely describes how to label such data with their likely causes[i]). This distinction is important because knowing how causes generate data allows a model to generate synthetic or 'simulated' data from given causes.

This generative description can be of how brain activity generates brain imaging data, or of how states in the world evolve and affect an agent's decision-making (eg, figure 1, described in the next section); in the latter case, commonly used in Computational Psychiatry, we are modelling the brain's own model of the world.[20] By altering key parameters in our generative models of agents' brains, we can observe what effects they have on decision-making and use this information to optimise experimental design or make counter-intuitive predictions. Bayesian statistics and machine learning techniques then allow this entire description to be tested against real data for goodness-of-fit. Comparisons of generative models by means of Bayesian model selection offer among the most rigorous and global comparative assessment of scientific hypotheses.[21]

Formulating a generative model requires an explicit description of the mathematical details of the cognitive or neural process from the outset. This is difficult but important, as it can force one not only to think hard about what particular constructs (such as 'attention' or 'salience') really mean,[22] but also to be very explicit about assumptions and ignorance.

Computational Psychiatry is not biologically reductionist

Computational Psychiatry is of course reductionist in the sense that it wishes to reduce cognitive processes to computations. Importantly, however, it does not view genes, neurotransmitters or neural circuits as causes of mental illness separate to the context in which the agent operates.[23] Indeed, it is precisely the agent's environment (both physical and social) which the nervous system may be trying to model, and which our models must also reflect.[8 19]

It is true that the new science of epigenetics (heralded by Engel[4] almost 40 years ago) also places genes firmly in their environmental context; but as an explanation of mental illness, a gene–environment interaction in the absence of any computational specification is a sandwich without the meat. What Psychiatry ultimately wants to know is: How and why does this gene–environment interaction change inference (and thereby experience and behaviour)?

Computational Psychiatry is not artificially categorical

One important, but unfulfilled, aspiration of the architects of DSM-5 was to move beyond purely categorical diagnoses to a more dimensional system, as it seems our current categories are not valid at the clinical[24] or genetic[25] levels. Such an approach would not classify a person with psychosis as just one of 'schizophrenic', 'bipolar' or 'schizoaffective', but might instead score them on scales of 'manic' and 'depressive' mood symptoms, 'positive' (delusional and hallucinatory) and 'negative' (avolitional) psychotic symptoms, and 'cognitive impairment'.

Computational Psychiatry can accommodate and inform both categorical and dimensional approaches—each driven by data. For example, one might find that depressed participants and controls differ continuously (dimensionally) on a certain parameter derived from a certain computational model (eg, 'reward prediction error signalling').[26] Alternatively, one might find evidence that different models are used by distinct groups (ie, possible categories) to perform the same task, for example, patients with schizophrenia with high or low negative symptoms,[27] or those with remitted psychosis and controls.[28] More generally, having defined alternative (eg, categorical vs dimensional) models, Computational Psychiatry allows one to assess the evidence for competing theories formally, for instance using Bayesian model comparison. Identifying computational categories and dimensions in this way ought to improve both psychiatric nosology[29] and the targeting and monitoring of treatments.

THEORETICAL FRAMEWORK

Computational theories of mind often mirror contemporaneous engineering practices, and vice versa. In the 1970s, for example, philosophers of mind took inspiration from the computer's manipulation of symbols according to deterministic syntactic rules, and sought to explain how humans might also think logically.[30] Recently, computer science has tried to make machines that can learn and make probabilistic inferences using uncertain or incomplete data, as biological agents can. In this section, we briefly introduce some necessary theoretical constructs in probabilistic inference and action selection before discussing some Computational Psychiatry approaches to depression and psychosis.

Inferring the present

Put as simply as possible, the brain's fundamental computational task is to infer the state of its environment and choose actions on that basis. Unfortunately, neither its sensory data nor its prior knowledge is completely reliable, and so the brain must use both sources of information—taking into account their uncertainty—to perform its task. The optimal combination of uncertain information is given by Bayes' theorem, in which a 'prior' (the initial expectation of the state of the environment) is combined with a 'likelihood' (the probability of the sensory input, given that expectation) to compute a 'posterior' (an updated estimation of the state of the environment). For simplicity, these probability distributions are often assumed

[i]In mathematical terms, discriminative models learn p(causes|data)—the probability of some causes, given the data—whereas generative models learn the reverse, p(data|causes), and use that (and p(causes) and Bayes' theorem) to compute p(causes|data).

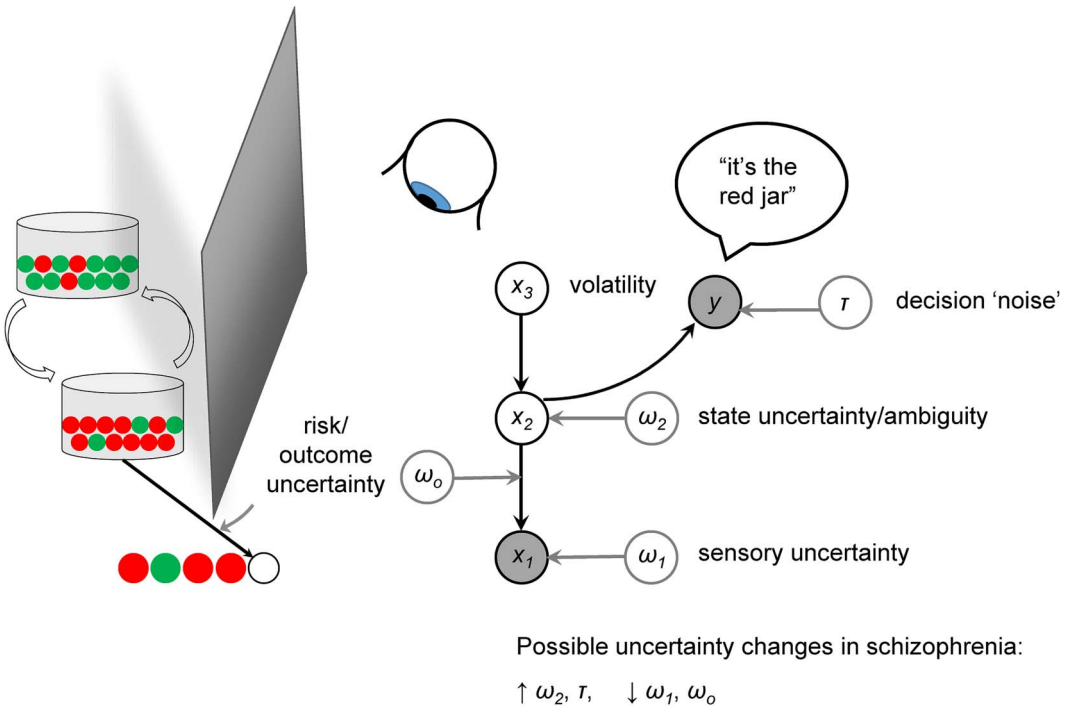

Figure 1 A hierarchical generative model, illustrated using the 'beads' or 'urn' task. On the left, two jars are hidden behind a screen, one containing mostly green and some red balls, the other the converse. A sequence of balls is being drawn from one of these jars, in view of an observer, who is asked to guess from which jar they are coming. We have illustrated a simple hierarchical generative model of this process on the right: the observer is using such a model to make his/her guess. Variables in shaded circles are observed, and variables in unshaded circles are 'hidden' (ie, part of the model only). At the bottom of the model is x_1, the colour of the currently observed bead. Uncertainty about this quantity (eg, if the light is low or if the participant is colour-blind) is denoted as ω_1. At the next level of the model is x_2, the belief about the identity of the current jar, and its associated uncertainty ω_2, known as state uncertainty or ambiguity. Another form of uncertainty, risk or outcome uncertainty (ω_0) governs the relationship between the identity of the jar and the next outcome: Even if we are sure of the jar's identity, we cannot be certain of the colour of the next bead. At the top of the model is the belief about the probability that the jars could be swapped at any time, known as volatility. We have not shown them here but this could have its own associated uncertainty, and there could be further levels above this. Last, the participant must use his/her belief about the identity of the jar to make a guess: The mapping between this belief and the response y is affected by a degree of stochasticity or decision 'noise', τ. In schizophrenia, there may be too much uncertainty (ie, lower precision) in higher hierarchical areas that encode states or make decisions, and an underestimation of uncertainty in lower (sensory) areas.

to be of a kind that can be represented by a few 'sufficient statistics'; for instance, the mean and precision (inverse variance) of a Normal distribution, in this case both prior and likelihood, can be conveniently weighted by their (scalar) precision.

Aside from their inherent uncertainty, the statistics of natural sensory stimulation are also extremely complex. Nevertheless, as a consequence of the hierarchical structure of the environment, they contain patterns: These patterns are easiest to interpret if the brain's prior beliefs respect the hierarchical structure in its sensory data—that is, if they take the form of a hierarchical model. Hierarchical models explain complex patterns of low-level data features in terms of more abstract causes: for example, the shape that describes a collection of pixels, or the climate that describes annual variation in weather. Hierarchical models are particularly important in the face of complex situations, both behavioural and sensory, and allow for highly efficient decompositions that greatly support planning and simplify optimal decision-making.[31–33]

Hierarchical generative models can use predictive coding (or other methods) to predict low-level data by exploiting their high-level descriptions, for example, reconstructing the missing part of an image.[34 35] In predictive coding, a unit at a given hierarchical level sends messages to one or more units at lower levels which predict their activity; discrepancies between these predictions and the actual input are then passed back up the hierarchy in the form of prediction errors. These prediction errors revise the higher level predictions, and this hierarchical message passing continues in an iterative fashion.

Exactly which predictions ought to be changed in order to explain away a given prediction error is a crucial question for hierarchical models. An approximately Bayesian solution to this problem is to make the biggest updates to the level whose uncertainty is greatest relative to the incoming data at the level below: that is, if you are very uncertain about your beliefs, but your source is very reliable, you ought to change your beliefs a lot.

Say, for example, I am walking at dusk, and I perceive the movement of a bush in my peripheral vision. I might explain this at various hierarchical levels, as (1) the bush did not actually move; it was a trick of the light; (2) the wind was moving the bush; (3) an animal was moving the bush and (4) a man hiding in the bush, intending to rob me, moved it. The conclusion that I draw will be determined by how precise my beliefs (at each level) are that (1) I saw movement; (2) the wind probably did not cause it; (3) there are probably no animals in the vicinity and (4) there is probably no mugger in the vicinity. The most uncertain (least precise) of these beliefs will have to change (assuming, for the sake of argument, that their likelihoods are equivalent), with very distinct consequences for my subsequent behaviour. Put more formally, the uncertainty (inverse precision) at each level helps determine the learning rate at that level, that is, the size of the adjustments that are made to explain new data.[34]

A classic psychology experiment illustrates uncertainty at different levels (see figure 1). Imagine you are shown two jars of beads, one containing 85% green and 15% red beads, the other 85% red and 15% green. The jars are then hidden and a sequence of beads is drawn (with replacement)—GGRGG. You are asked to guess the colour of the next bead. Even if you are quite certain of the identity of the jar (say, green), you will still be only 85% certain that the next bead will be green. This is 'outcome uncertainty' or risk. Imagine you see more beads—the total sequence is GGRGGRR. Now you are very uncertain about the identity of the jar. This is 'state uncertainty' or ambiguity. Imagine you see a much longer sequence—GGRGGRRRRRGGGRGGGGGGRGGGG. From this, it seems that the jar changes from green (5 draws), to red (5 draws), to green (remaining sequence). Such temporal changes in hidden causes give rise to 'volatility';[36 37] for example, someone surreptitiously switching the jar during the experiment.

Now suppose that although the real proportions are 85% and 15%, a malicious experimenter misleadingly told you that they are 99.9% and 0.1%. From the 25 draw sequence above, you might reasonably conclude that the jars had actually changed eight times—whenever the colour changed. This is what happens when the precision at the bottom of a hierarchical model is too high relative to the precision at the top: Following a sensory prediction error, the model concludes that there must have been a change in the environment (in this simplistic example, the jar), rather than 'putting it down to chance'.

This precision imbalance might contribute to the well-known 'jumping to conclusions' reasoning bias in schizophrenia[38] (although another cause might be noisy decision-making[28]) and the formation of delusional beliefs themselves, which commonly arise in an atmosphere of vivid sensory experiences and strange coincidences.[14] We return to the subject of delusions in the Psychosis section.

Weighing the future

On top of the inference about current stimuli and states, the brain needs to solve a second, orthogonal and complicating problem, which arises from the fact that behaviours have both immediate and future consequences. A brief moment of pleasure can have nasty consequences and, though tempting, might best be avoided. Conversely, pain now might result in even greater pleasures later. Optimal behaviour needs to weigh the future against the present. Even in the rare circumstances where inference about the present is perfectly realisable, the brain therefore faces a second set of uncertainties. As the future is not known, the values of possible actions (their summed future rewards and punishments) have to be somehow estimated.

The field of reinforcement learning (RL) has delineated two fundamentally different ways in which past experience is used to estimate and predict future rewards and punishments: so-called model-based (MB) and model-free (MF) cognition.

In MB or goal-directed cognition, experience is compiled into a (possibly hierarchical) generative model of the world—a mechanistic, causal understanding of the causes and consequences of actions and events. When faced with a particular situation, this model can be searched, and the quality of various behaviours deduced—even if they have never been tried or experienced. As this involves somehow simulating or inferring future possibilities, it can have high computational costs.

In MF or habitual cognition, conversely, the agent does not store information about state transitions (ie, exactly what is likely to happen next if a particular action is performed); instead, the agent merely records how much reinforcement is obtained when a certain state s_t is visited at time t or an action a_t is taken. The agent then computes the discrepancy between the expected outcome $V_t(s_t)$ and the actually obtained outcome r_t—the prediction error $\delta_t = r_t + V_t(s_{t+1}) - V(s_t)$. MF learning adds a fraction ε of the reward prediction errors δ_t to the expectations every time the state s_t is visited: $V_{t+1}(s_t) = V_t(s_t) + \varepsilon\delta_t$ and thereby reduces the discrepancy between expected and received reinforcements. Under some conditions, this will reveal the true, consistent set of values that incorporate 'future' outcomes to the extent that these have followed past choices.

MF learning is computationally undemanding but slow and inflexible, so it is undermined by sudden changes to the environment or to the valuation of rewards. One example of this is a rat learning how to obtain salty food in a maze: MB learning would create knowledge of the internal structure of the maze and the kind of rewards available within it, whereas MF learning would register a given sequence of left/right turns as being the best thing to do. If the rat's usual route were blocked, or if it were thirsty, then the MF information would not be useful. The MF account can be extended—for example, using hierarchical models—to allow more sensitivity to changes.[39]

We next describe these issues in the context of depression and schizophrenia, with the discussion of depression focusing more on valuation, that is, weighing the future, and that of schizophrenia on inference in the present.

VALUATION IN DEPRESSION

Depression is by its nature an aversive state, usually accompanied by negative thoughts about the self, the world and the future, and sometimes characterised by reduced interest and/or pleasure known as anhedonia. However, the core symptoms of depression—sadness, a lack of energy and a reduced ability to enjoy things—are also a frequent but temporary feature of everyday affective experience. The symptoms can become a condition of potentially life-threatening severity if they become part of a vicious circle of negative affect, cognition and behaviour that is impervious to positive influences.

Computational descriptions of normative and resource-rational choices—RL and Bayesian decision theory—reveal many ways in which such a vicious circle could arise. We review how a stable state of anhedonia might exist in RL, and then close with a few speculative suggestions about the most likely paths to this state. Parts of this are described in more detail elsewhere.[40 41]

Primary utility

With respect to the potential causes of depression, the most obvious candidate in RL is the so-called utility or reward function r. This function uses one scalar number to describe how rewarding (increasingly positive) or punishing (increasingly negative) events in the world are. Some kinds of events may have genetically encoded reward or punishment utilities—most likely, only those of direct relevance to the individual's genetic fitness. Other events acquire 'utility' through experience and inference. A very broad undervaluation as in anhedonia could arise from reductions in hard-wired primary utility functions. However, evidence for this is both scant and complex. The apparent utilities of two likely primary events—pleasure derived from sweet tastes and pain—are not reliably blunted in depression when measured in the laboratory.[42 43] This contrasts with richer and more complex stimuli such as pictures of facial expressions, movies and music, where blunted affective ratings and physiological responses are reliably present in the appetitive and aversive domains.[44] Since the affective value of these more

complex stimuli has to be constructed and inferred, unlike that of events with primary utility, this points to the impaired inference about value as the central driver of undervaluation in depression.

Inferred value in depression

RL is concerned with the problem of assigning and inferring value. Specifically, it provides a set of techniques to infer the long-term primary utility, either when occupying any one particular state or when faced with a particular stimulus devoid of any intrinsic primary utility itself, such as a picture. Either of the two classes of approaches to valuation, MB and MF, could in principle underlie aberrant valuation of complex stimuli.

We first briefly examine MF valuation, which, as introduced briefly above, depends on reward prediction errors. Three types of experimental set-up in humans speak to this. First are experiments examining prediction errors per se without any requirement for learning (ie, the contingencies are fully instructed), such as the monetary incentive delay task.[45] In general, such tasks have not yielded consistent differences between depressed and control subjects in either behavioural measures or their neural correlates in the areas most strongly associated with MF learning, such as the ventral striatum.[46–50] Second are studies explicitly examining the kind of trial-by-trial learning described by MF valuation. While these studies have shown slightly more consistent group differences at the neural level, for instance in the ventral striatum, their interpretation is complicated both by variable results in the midbrain dopaminergic regions and often by the absence of any behavioural effects.[26 51–54] Third are studies using a probabilistic response bias task.[55] RL modelling of this task suggested that anhedonia was not related to MF learning.[56]

In contrast, several features of MB valuation appear to be involved in depression. Cognitive theories of depression have long emphasised the importance of schemas,[57 58] which, when activated by environmental triggers, lead to negative automatic thoughts and consequent aversive feelings. From a Bayesian perspective, these schemas can be viewed as priors, and the automatic thoughts as the result of a combination of the priors with the triggering sensory events. One example of this is learned helplessness, a model of depression in which healthy animals are exposed to uncontrollable stressors and come to show a variety of depression-like behavioural anomalies in other situations.[59] These effects can arise from prior beliefs about the achievability of desirable outcomes.[60] One would expect that, being part of the model of the world, the prior would act within the goal-directed MB valuation system. Indeed, the behavioural effects of uncontrollable stressors depend on midline prefrontal areas[61] known to be involved in MB reasoning.[62]

One feature that is of particular interest is emotion regulation. We suggest that it again arises chiefly in MB evaluation and might be understood in terms of meta-reasoning. Since most valuation problems are computationally demanding, the MB evaluator faces the meta-reasoning problem of how to allocate its resources efficiently—that is, how to choose which evaluative (internal) actions maximise the chances of choosing the best (external) action. For example, one could sacrifice an exhaustive search of all possible outcomes to save time by only evaluating a small number of more likely scenarios. This problem has many of the features of the original valuation problem, but differs in that in theory it only incurs computational costs and not those of the real world (eg, pain). In practice, however, this distinction does not seem to hold for humans. Since imagination of aversive events has emotionally aversive consequences, internal simulations themselves also incur some of the same costs as real-world experience. Indeed, it has been found that MB valuation is exquisitely sensitive to simulated events. Healthy subjects robustly avoid plans that involve losses.[63] It appears that patients with depression have a deficit in this inhibition of aversive processing,[64 65] with aversive stimuli hijacking rather than inhibiting processing:[66] This is a potential cause of the repetitive negative thoughts typical of rumination[67]—a key component of the depressive vicious circle.

PRECISION AND D_2 RECEPTORS IN PSYCHOSIS

Having discussed aspects of valuation in depression, we now turn to a discussion of inference in schizophrenia. Specifically, we explore how various neurobiological abnormalities in schizophrenia might be characterised in computational terms, and how these characterisations might aid our understanding of the disorder. We discuss reductions in synaptic gain in higher hierarchical areas, and increased presynaptic dopaminergic availability and its consequences for tonic and phasic dopaminergic signalling in the striatum.

Psychosis, synaptic gain and precision

What are the main cortical abnormalities in schizophrenia and what do they have in common (reviewed in detail elsewhere[68])? One key abnormality is thought to be hypofunction of the *N*-methyl-D-aspartate receptor (NMDA-R)—a glutamate receptor with profound effects on synaptic gain (due to its prolonged opening time) and synaptic plasticity (via long-term potentiation or depression)—in both the prefrontal cortex (PFC) and hippocampus (HC). A second is the reduced synthesis of γ-aminobutyric acid (GABA) by inhibitory interneurons in PFC. A third is the hypoactivation of D_1 receptors in PFC (we shall discuss striatal hyperactivation of D_2 receptors in the next section).

These abnormalities could all reduce synaptic gain in PFC or HC, that is, around the top of the cortical hierarchy. Synaptic gain (or 'short term' synaptic plasticity[69]) refers to a multiplicative change in the influence of presynaptic input on postsynaptic responses. NMDA-R hypofunction and D_1 receptor hypoactivity are most easily related to a change in synaptic gain. Similarly, a GABAergic deficit might cause a loss of 'synchronous' gain. Sustained oscillations in neuronal populations are facilitated through their rhythmic inhibition by GABAergic interneurons, putatively increasing communication between neurons that oscillate in phase with each other.[70]

How can synaptic gain (and its loss) be understood in computational terms? One answer rests on the idea that the brain approximates and simplifies Bayesian inference by using probability distributions that can be encoded by a few 'sufficient statistics', for example, the mean and its precision (or inverse variance). While precision determines the influence one piece of information has over another in Bayesian inference, synaptic gain determines the influence one neural population has over another in neural message passing. The neurobiological substrate of precision could therefore be synaptic gain,[22] and a loss of synaptic gain in a given area could reduce the precision of information encoded there.

A loss of synaptic gain in PFC or HC would diminish their influence over lower level areas. In the model, this would correspond to a loss of influence (ie, precision) of the model's more abstract priors over the more concrete sensory data. To the extent to which the higher levels extract and represent more stable, general features of the world, their loss might make the

world look less predictable and more surprising. This simple computational change can describe a great variety of phenomena in schizophrenia (figure 2; more references and simulations of some of these phenomena are elsewhere[68]):

- At a neurophysiological level, responses to predictable stimuli resemble responses to unpredicted stimuli, and vice versa, in both perceptual electrophysiology experiments (eg, the P50 or P300 responses to tones[71]) and cognitive functional MRI paradigms;
- At a network level, higher regions of the cortex (ie, PFC and HC) have diminished connectivity to the thalamus relative to controls, whereas primary sensory areas are coupled more strongly with this region;[72]
- At a perceptual level, a greater resistance to visual illusions[73] (which exploit the effects of visual priors on ambiguous images, eg, the famous 'hollow-mask' illusion[74]) and a failure to attenuate the sensory consequences of one's own actions, which could diminish one's sense of agency;[75]
- At a behavioural level, impaired smooth visual pursuit of a predictably moving target, but better tracking of a sudden unpredictable change in a target's motion.[76]

An alternative interpretation of these changes is that pathology in the PFC or HC (eg, in postsynaptic signalling and neurotrophic pathways[77]) might impair the formation and representation of prior beliefs more generally, rather than directly and selectively affecting only a separate representation of their certainty. Indeed, if this were the case, then reducing the influence of aberrant beliefs on sensory processing might even be computationally adaptive (and physiological rather than pathophysiological), since it would reduce their (possibly misleading) influence on inference.

How do the above ideas relate to the actual symptoms of psychosis? A reasonable hypothesis would be that a loss of high-level precision in the brain's hierarchical inference might result in diffuse, generalised cognitive problems (as are routinely found in schizophrenia[78]) and overattention to sensory stimuli (as is found in the 'delusional mood'; similar in some respects to the loss of central coherence in autism—see below). In addition, it could lead to the formation of more specific unusual beliefs, as the reduced high-level precision permits updates to beliefs that are larger and less constrained. This is because the precision of (low-level) prediction errors is much higher, relative to the (high-level) prior beliefs (an imbalance reflected in connectivity analyses[72]). However, one might expect that these unusual beliefs should be fleeting—as they themselves would be vulnerable to rapid updating—unlike delusions.

This account raises two important (and, we propose, related) questions that we address in the next section. First, if high-level precision is generally low, why do delusions, which appear to exist at a reasonably high (conceptual) level in the hierarchy, become so fixed? Second, what is the computational impact of the best-established neurobiological abnormality in schizophrenia—an elevation in presynaptic dopamine?[79]

Striatal presynaptic dopamine elevation

The positive symptoms of schizophrenia are strongly associated with the elevated presynaptic availability of dopamine in the dorsal (associative) striatum,[79] and are reduced by D_2 receptor (but not D_1 receptor) antagonists (although neither is always the case[80]). Increased stimulation of striatal D_2 receptors might then be a sufficient cause of psychosis, but the exact nature of this stimulation and how it causes psychotic symptoms remains unclear.

In electrophysiological studies, dopamine neurons show both tonic and phasic firing patterns;[81] D_1 receptors are most sensitive to phasic bursts, whereas tonic activity and phasic pauses are best detected by D_2 receptors.[82] These patterns cannot be distinguished using brain imaging in humans, unless their computational roles can be modelled and thus their quantities inferred from behaviour. It is as yet unclear how the increased presynaptic availability of dopamine alters these patterns in schizophrenia: One might expect that both tonic and phasic release would increase in proportion, but an increase in tonic release could reduce phasic release, for example, via inhibitory presynaptic receptors,[83] and these two modes of release have also been argued to be at least partially independent.[81] We now explore how these patterns may be disrupted in schizophrenia, and how this might affect computations.

Tonic dopamine signalling

Tonic striatal D_2 hyperstimulation is thought to increase inhibition of the corticostriatothalamocortical loops in the so-called indirect pathway through the striatum. The indirect pathway itself contains two inhibitory pathways. One (via the subthalamic nucleus) causes blanket inhibition of action and acts as a brake, but the other is channelised[84] such that it can help switching to alternative actions.

If the indirect pathway enables switching, then increased tonic D_2 receptor activity in the striatum ought to oppose this (interestingly, D_2 receptors also suppress alternative task 'rules' in PFC[85]). Indeed, reversal learning performance decreases with increasing occupancy of D_2 receptors in the dorsal[86 87] and ventral[81] striatum, and with genetic variants in the dopamine transporter that might increase tonic dopamine.[88] Sufferers of schizophrenia are impaired at reversal learning over and above their generalised cognitive impairment,[89 90] which is in keeping with these findings. D_2-mediated inflexibility might even make delusions so resistant to change.

From a computational point of view, this striatal D_2 hyperstimulation could reduce an agent's perception of volatility in the world (x_3 in figure 1), causing action tendencies (known as policies) to become more fixed and have (incorrectly) high 'precision'. Indeed, work in addiction has similarly argued that dopamine promotes rather less flexible habits over goal-directed choices.[91] This is interesting because it is conceivable that a hyperdopaminergic increase in (dorsal striatal) precision of policies might occur as an adaptation to a loss of (prefrontal) high-level precision (ω_2 in figure 1),[68] that is, that the excessive dorsal striatal dopamine release found in prefrontal dysfunction[92 93] is an attempt to stabilise thoughts and action selection in the face of cognitive instability.

Even if this is so, one must still ask why so many other risk factors for schizophrenia—for example, social isolation or subordination, prenatal or perinatal adversity, and acute stress—cause dopamine hyperactivity?[79] One could argue that the computational commonality among these factors is an increase in predicted environmental volatility, but raised tonic dopamine release makes decisions less, not more, volatile.

Phasic dopamine signalling

The phasic responses of dopamine neurons comprise bursts and pauses, which facilitate the excitatory 'direct' pathway through excitatory D_1 receptors, and facilitate the inhibitory 'indirect' pathway through inhibitory D_2 receptors, respectively. These have been proposed to reflect reward prediction errors[94] (but also aversive prediction errors[95] and the precision or 'salience' of prediction errors[96]) in the ventral and dorsal striatum.

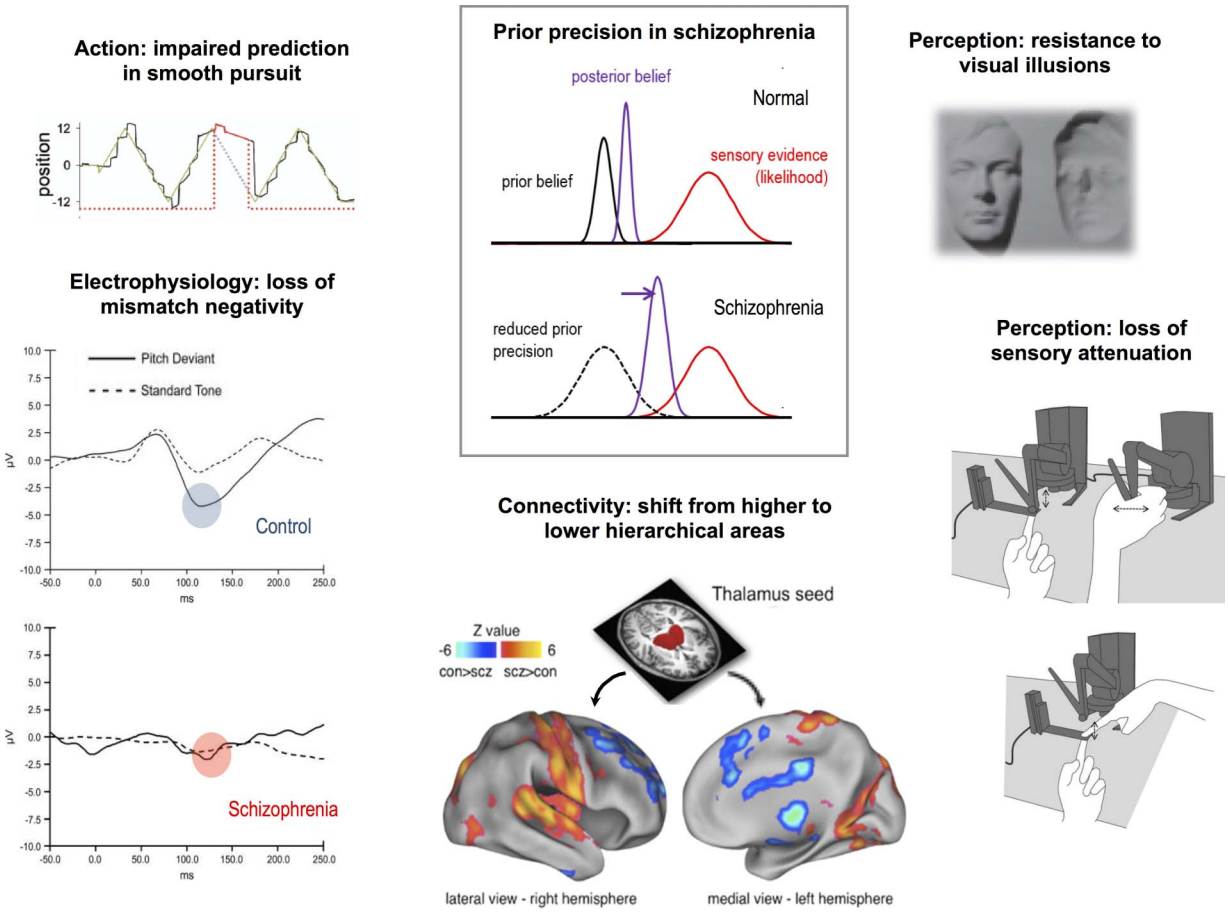

Figure 2 Effects of a hierarchical precision imbalance in schizophrenia. A loss of precision encoding in higher hierarchical areas would bias inference away from prior beliefs and towards sensory evidence (the likelihood), illustrated schematically in the middle panel. This single change could manifest in many ways (moving anticlockwise from left to right). (i) A loss of the ability to smoothly pursue a target moving predictably (in this plot, the patient with schizophrenia constantly falls behind the target in his eye tracking, and has to saccade to catch up again); when the target is briefly stabilised on his retina (to reveal the purely predictive element of pursuit), shown as the red unbroken line, his/her eye velocity drops very significantly (figure adapted from Hong *et al*[76]). (ii) These graphs illustrate averaged electrophysiological responses in a mismatch negativity paradigm, in which a series of identical tones is followed by a deviant (oddball) tone; in the control subject, the oddball causes a pronounced negative deflection at around 120 ms (blue circle), but in a patient with schizophrenia, there is no such deflection (red circle); that is, the brain responses to predictable and unpredictable stimuli are very similar (figure adapted from Turetsky *et al*[71]). (iii) The physiological change underlying the precision imbalance is a relative decrease in synaptic gain in high hierarchical areas, and a relative increase in lower hierarchical areas. This change would also manifest as an alteration in connectivity, shown here as significant whole brain differences in connectivity with a thalamic seed between controls and patients with schizophrenia; red/yellow areas are more strongly coupled in those with schizophrenia, and include primary sensory areas (auditory, visual, motor and somatosensory); blue areas are more weakly coupled, and include higher hierarchical areas (medial and lateral prefrontal cortex, cingulate cortex and hippocampus) and the striatum (figure adapted from Anticevic *et al*[72]). (iv) An imbalance in hierarchical precision may lead to a failure to attenuate the sensory consequences of one's own actions,[75] here illustrated by the force-matching paradigm used to measure this effect. In this paradigm, the participant must match a target force by either pressing on a bar with their finger (below) or using a mechanical transducer (top): Control subjects tend to exert more force than necessary in the former condition, but patient with schizophrenia do not (figure adapted from Parees *et al*[119]). (v) A loss of the precision of prior beliefs can cause a resistance to visual illusions that rely on those prior beliefs for their perceptual effects. Control subjects perceive the face on the right as a convex face lit from below, due to a powerful prior belief that faces are convex, whereas patients with schizophrenia tend to perceive it veridically as a concave (hollow) face lit from above.

Actor-critic models[6] propose that reward prediction errors act as signals which teach the ventral striatum (critic) the values of states, and the dorsal striatum (actor) to associate states with optimal actions (by increasing the excitation or reducing the inhibition of the currently selected action[6]), with recent evidence for a causal role of dopamine in this regard.[97]

In fMRI paradigms designed to elicit reward prediction error and reward prediction signals, patients with schizophrenia show diminished appropriate activations[98] but greater inappropriate activations in the ventral striatum,[99] compared with controls. Similar patterns were observed in an associative learning task without explicit rewards;[100] however, fMRI cannot tell whether these abnormalities are due to abnormal phasic dopamine signalling. Although behavioural responses in such tasks are often less abnormal than the underlying neural activity,[101] computational modelling of behaviour in the beads task[102] and reward learning tasks[27] suggests that the impact of phasic positive feedback is diminished in schizophrenia. This may indicate that an elevated tonic dopamine level is reducing phasic bursts (reward learning) but not phasic pauses (punishment learning), and indeed PET studies suggest an inverse relationship between tonic DA and phasic BOLD signals.[103] An alternative suggestion is that this apparent reward learning deficit may actually be caused by reduced working memory capacity.[104]

Incentive (and aberrant) salience

The theory of incentive salience proposes that ventral striatal dopamine signalling (whether phasic or tonic) gives motivational impetus to act on stimuli whose values have already been learnt.[105] Incentive salience is closely related to MF learning of values,[91 106] and might also be related to the precision or confidence of beliefs that actions will have preferred outcomes.[96] In the 'aberrant salience' hypothesis, Kapur[107] proposed that there is aberrant (ie, increased inappropriate) signalling of incentive salience in patients with positive psychotic symptoms. Unmedicated prodromal psychotic patients do experience—in proportion to their positive symptoms—irrelevant features of stimuli in a reward learning task as 'aberrantly salient' (although this is not obviously reflected in their reaction times), but their striatal activations are harder to interpret.[108]

A weakness of the aberrant salience hypothesis is that the connection between aberrant motivational signalling and the abnormal inference (hallucinations) and abnormal learning (delusions) found in positive symptoms is not intuitive (assuming that delusions do indeed involve abnormal learning). One could also argue that aberrant motivational salience works best as an account of manic psychosis—in which the patient is energised and perceives events in a positive light—rather than schizophrenic psychosis, which is often aversive in nature. Conversely, diminished appropriate salience signalling (not part of the original hypothesis but identified in several studies) provides a plausible explanation for negative symptoms; and, indeed, a loss of ventral striatal activation to rewards has been shown to be proportional to negative symptoms in unmedicated patients with schizophrenia.[109]

Aside from aberrant salience, there are many other potential explanations for negative symptoms;[15] for example, pronounced asymmetry in learning (ie, a failure to learn stimulus-reward associations but intact learning of stimulus-punishment associations), a failure to infer the values of actions (cf. anhedonia in depression), greater discounting of rewards that require effort,[110] and a loss of uncertainty-driven exploration such that valuable states are never discovered.

CONCLUSION

In this brief overview, we have attempted to highlight some aspects of the Computational Psychiatry approach to characterising and measuring the brain's inferences. We have not had space to review Computational Psychiatry approaches to many mental disorders, such as anxiety,[111] personality disorder,[8] autism,[112] attention deficit hyperactivity disorder,[113] addiction,[6] functional symptoms[114] and others. We focused instead on two examples: The first was that of depression, where we suggested that MB valuation may be the at the root of anhedonia.[27] Our second example was the concept of reduced high-level precision in schizophrenia. We must emphasise, however, that their description in the context of specific disorders should not necessarily be taken to imply specificity to these disorders. The dysfunctions in meta-reasoning described here in depression could probably be applied equally well to certain anxiety disorders.[115] Anhedonia is also found in schizophrenia, and in this context just as in depression, hedonic responses to primary rewards seem normal,[116] yet pleasure-seeking behaviour is reduced. If so, Computational Psychiatry might help identify transdiagnostic computational mechanisms: we must then investigate the extent to which such mechanisms share biological substrates.

Similarly, alterations in the use or representation of uncertainty may underlie phenomena not just in schizophrenia, but also in autism (eg, a resistance to visual illusions, and sensory overattention[112]). However, an important difference between the disorders may be that this reduction in precision arises earlier in development in autism. Thus, while sufferers of both disorders may have uncertainty about the mental states of others, those with autism have never learnt to attribute mental states to others. In contrast, those with schizophrenia have done so in the past, but in the present they find themselves in a state of high anxiety and uncertain of others' intentions, possibly suggesting why paranoid persecutory ideas might be commonly found in schizophrenia, but not autism. In this case, then, common computational mechanisms may undergo distinct interactions with the environment. In both cases, the hope is that Computational Psychiatry will identify patterns that map more closely onto the underlying neurobiology than current diagnoses do.

Aside from disorders that are thought of as purely psychiatric, Computational Psychiatry has been used to examine putatively dopaminergic computations in neurological conditions such as Parkinson's disease and Tourette's syndrome. In some classic studies, Frank and colleagues devised a probabilistic task which quantifies individuals' abilities to learn from positive and negative feedback, encoded by dopamine bursts and pauses, respectively. They showed that dopamine-depleted patients with Parkinson's disease were better than controls at learning from negative feedback, but dopaminergic medication reverses this bias.[6]

This asymmetry could underlie the phenomenon of pathological gambling in the context of dopamine agonist treatment: such patients may be able to learn only from their wins but not their losses. Interestingly, the opposite asymmetry (better learning from positive feedback, reversed by antidopaminergic medication) was seen in the hyperdopaminergic Tourette's syndrome.[117] Computational Psychiatry approaches are also being used to investigate motivation and effort cost in apathy[118]—found in Parkinson's disease and numerous other neurological conditions—and it will be interesting to see how many underlying computational mechanisms are shared with negative symptoms in schizophrenia.

Computational Psychiatry is no panacea, however, and several important pitfalls ought to be mentioned. The first is the interpretation of the results of Bayesian model selection: the 'best' model is extremely unlikely to model the true generative process correctly, and may be but the best of a set of bad models. Model comparison must always be accompanied by model validation, which involves generating surrogate data from the model and comparing it qualitatively with the data of interest.

Generating data will in fact often identify systematic failures of models that may confound the interpretation of their parameters. This will require the addition of ad hoc parameters to explain specific aspects of the data that may interfere with inference, but not be of interest. Consider a data set where a patient has an idiosyncratic preference for responding to the left or to the right. If such a preference is not allowed for in the model, then the parameters of the model will be forced to explain such a preference. This may lead to spurious conclusions, just as the failure to recognise such irrelevant idiosyncrasies may confound classical analyses.

Furthermore, modelling does not replace careful experimental design. If a task does not exploit a particular computation, then that computation cannot be examined simply by fitting a model requiring that computation to the data. Put differently, the complexity of the model must be supported by complex data, and this in turn requires appropriate experimental design. The last pitfall we shall mention concerns a ubiquitous parameter in

decision-making models: the 'temperature' parameter of the softmax action selection function. This controls the stochasticity of decision-making, such that the greater it is, the less decisions reflect the values of different options, that is, the more random they are. It is always possible, however, that a difference in the stochasticity of responses between groups reflects a non-random process that has not been included in the model. Hence, it is important to examine whether the noise assumed by the model actually matches that observed in the data.

Our fundamental message is that thinking of the brain as having to solve inferential problems can be a fruitful way of generating testable computational hypotheses about psychiatric disorders. The Bayesian perspective formalises the key aspects of inference and underlines the importance of uncertainty, while the RL view formalises the key aspects of choice. Characterising psychiatric disorders as problems of inference or learning—whether in the domain of rewards, threats, somatic percepts, 'external' percepts or social inferences—makes them tractable to analysis with these techniques.

Acknowledgements The authors would like to thank Professor Peter Dayan, Professor Karl Friston and our reviewers for their valuable comments on the manuscript.

Funding RAA is supported by the National Institute of Health Research. QJMH is supported by a Swiss National Science Foundation grant (320030L_153449/1). JPR is supported by the Wellcome Trust (101798/Z/13/Z).

Competing interests None declared.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES

1 American Psychiatric Association. *DSM 5*. American Psychiatric Association, 2013.
2 World Health Organization. *ICD-10: International classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Geneva: WHO, 1992.
3 Kirmayer LJ, Crafa D. What kind of science for psychiatry? *Front Hum Neurosci* 2014;8:435.
4 Engel GL. The need for a new medical model: a challenge for biomedicine. *Science* 1977;196:129–36.
5 Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* 2013;11:126.
6 Maia TV, Frank MJ. From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci* 2011;14:154–62.
7 Huys QJ, Moutoussis M, Williams J. Are computational models of any use to psychiatry? *Neural Netw* 2011;24:544–51.
8 Montague PR, Dolan RJ, Friston KJ, *et al*. Computational psychiatry. *Trends Cogn Sci (Regul Ed)* 2012;16:72–80.
9 Deserno L, Boehme R, Heinz A, *et al*. Reinforcement learning and dopamine in schizophrenia: dimensions of symptoms or specific features of a disease group? *Front Psychiatry* 2013;4:172.
10 Huys QJM. Computational psychiatry. In: Jaeger D, Jung R, eds. *Encyclopedia of computational neuroscience*. Heidelberg: Springer Verlag, 2013. http://www.springer.com/biomed/neuroscience/book/978-1-4614-6674-1 (accessed 8 Sep 2014).
11 Stephan KE, Mathys C. Computational approaches to psychiatry. *Curr Opin Neurobiol* 2014;25:85–92.
12 Wang XJ, Krystal JH. Computational psychiatry. *Neuron* 2014;84:638–54.
13 Friston KJ, Stephan KE, Montague R, *et al*. Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry* 2014;1:148–58.
14 Corlett PR, Fletcher PC. Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *Lancet Psychiatry*. 2014;1:399–402.
15 Strauss GP, Waltz JA, Gold JM. A review of reward processing and motivational impairment in schizophrenia. *Schizophr Bull* 2014;40(Suppl 2):S107–16.
16 Hoffman RE, McGlashan TH. Neural network models of schizophrenia. *Neuroscientist* 2001;7:441–54.
17 Cohen JD, Braver TS, O'Reilly RC. A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges. *Philos Trans R Soc Lond B Biol Sci* 1996;351:1515–27.
18 Marr D. *A computational investigation into the human representation and processing of visual information*. WH San Francisco: Freeman and Company. Published Online First: 1982. http://www.contrib.andrew.cmu.edu/~kk3n/80-300/marr2.pdf (accessed 24 Feb 2014).
19 Moutoussis M, Trujillo-Barreto NJ, El-Deredy W, *et al*. A formal model of interpersonal inference. *Front Hum Neurosci* 2014;8:160.
20 Adams RA, Aponte E, Marshall L, *et al*. Active inference and oculomotor pursuit: the dynamic causal modelling of eye movements. *J Neurosci Methods* 2015;242:1–14.
21 Stephan KE, Penny WD, Daunizeau J, *et al*. Bayesian model selection for group studies. *Neuroimage* 2009;46:1004–17.
22 Feldman H, Friston KJ. Attention, uncertainty, and free-energy. *Front Hum Neurosci* 2010;4:215.
23 Read J, Bentall RP, Fosse R. Time to abandon the bio-bio-bio model of psychosis: exploring the epigenetic and psychological mechanisms by which adverse life events lead to psychotic symptoms. *Epidemiol Psichiatr Soc* 2009;18:299–310.
24 Van Os J, Gilvarry C, Bale R, *et al*. A comparison of the utility of dimensional and categorical representations of psychosis. UK700 Group. *Psychol Med* 1999;29:595–606.
25 Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013;381:1371–9.
26 Kumar P, Waiter G, Ahearn T, *et al*. Abnormal temporal difference reward-learning signals in major depression. *Brain* 2008;131:2084–93.
27 Gold JM, Waltz JA, Matveeva TM, *et al*. Negative symptoms and the failure to represent the expected reward value of actions: behavioral and computational modeling evidence. *Arch Gen Psychiatry* 2012;69:129–38.
28 Moutoussis M, Bentall RP, El-Deredy W, *et al*. Bayesian modelling of jumping-to-conclusions bias in delusional patients. *Cogn Neuropsychiatry* 2011;16:422–47.
29 Brodersen KH, Schofield TM, Leff AP, *et al*. Generative embedding for model-based classification of fMRI data. *PLoS Comput Biol* 2011;7:e1002079.
30 Fodor JA. *The language of thought*. Harvard University Press, 1975.
31 Botvinick MM, Niv Y, Barto AC. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* 2009;113:262–80.
32 Diuk C, Schapiro A, Natalia Cordova JR-F, *et al*. Divide and conquer: hierarchical reinforcement learning and task decomposition in humans. In: Baldassarre G, Mirolli M, eds. *Computational and robotic models of the hierarchical organization of behavior*. Springer Berlin Heidelberg, 2013:271–91.
33 Huys QJ, Lally N, Faulkner P, *et al*. The interplay of approximate planning strategies. *Proc Natl Acad Sci USA* 2015;112:3098–103.
34 Mathys C, Daunizeau J, Friston KJ, *et al*. A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* 2011;5:39.
35 Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 1999;2:79–87.
36 Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. *Neuron* 2005;46:681–92.
37 Payzan-LeNestour E, Bossaerts P. Risk, unexpected uncertainty, and estimation uncertainty: bayesian learning in unstable settings. *PLoS Comput Biol* 2011;7:e1001048.
38 Fine C, Gardner M, Craigie J, *et al*. Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cogn Neuropsychiatry* 2007;12:46–77.
39 Dayan P, Kakade S, Montague PR. Learning and selective attention. *Nat Neurosci* 2000;3(Suppl):1218–23.
40 Huys QJ, Daw ND, Dayan P. Depression: a decision-theoretic analysis. *Annu Rev Neurosci*. Published Online First: 2015.
41 Huys QJ, Guitart-Masip M, Dolan RJ, *et al*. Decision-theoretic psychiatry. *Clin Psychol Sci* 2015;3:400–21.
42 Dichter GS, Smoski MJ, Kampov-Polevoy AB, *et al*. Unipolar depression does not moderate responses to the sweet taste test. *Depress Anxiety* 2010;27:859–63.
43 Lautenbacher S, Spernal J, Schreiber W, *et al*. Relationship between clinical pain complaints and pain sensitivity in patients with depression and panic disorder. *Psychosom Med* 1999;61:822–7.
44 Bylsma LM, Morris BH, Rottenberg J. A meta-analysis of emotional reactivity in major depressive disorder. *Clin Psychol Rev* 2008;28:676–91.
45 Knutson B, Bhanji JP, Cooney RE, *et al*. Neural responses to monetary incentives in major depression. *Biol Psychiatry* 2008;63:686–92.
46 Forbes EE, Christopher May J, Siegle GJ, *et al*. Reward-related decision-making in pediatric major depressive disorder: an fMRI study. *J Child Psychol Psychiatry* 2006;47:1031–40.
47 Forbes EE, Hariri AR, Martin SL, *et al*. Altered striatal activation predicting real-world positive affect in adolescent major depressive disorder. *Am J Psychiatry* 2009;166:64–73.

48 Pizzagalli DA, Holmes AJ, Dillon DG, *et al*. Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *Am J Psychiatry* 2009;166:702–10.
49 Smoski MJ, Felder J, Bizzell J, *et al*. fMRI of alterations in reward selection, anticipation, and feedback in major depressive disorder. *J Affect Disord* 2009;118:69–78.
50 Zhang WN, Chang SH, Guo LY, *et al*. The neural correlates of reward-related processing in major depressive disorder: a meta-analysis of functional magnetic resonance imaging studies. *J Affect Disord* 2013;151:531–9.
51 Gradin VB, Kumar P, Waiter G, *et al*. Expected value and prediction error abnormalities in depression and schizophrenia. *Brain* 2011;134:1751–64.
52 Remijnse PL, Nielen MM, van Balkom AJLM, *et al*. Differential frontal-striatal and paralimbic activity during reversal learning in major depressive disorder and obsessive-compulsive disorder. *Psychol Med* 2009;39:1503–18.
53 Robinson OJ, Cools R, Carlisi CO, *et al*. Ventral striatum response during reward and punishment reversal learning in unmedicated major depressive disorder. *Am J Psychiatry* 2012;169:152–9.
54 Steele JD, Kumar P, Ebmeier KP. Blunted response to feedback information in depressive illness. *Brain* 2007;130:2367–74.
55 Pizzagalli DA, Jahn AL, O'Shea JP. Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biol Psychiatry* 2005;57:319–27.
56 Huys QJ, Pizzagalli DA, Bogdan R, *et al*. Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biol Mood Anxiety Disord* 2013;3:12.
57 Beck AT. *Depression: clinical, experimental, and theoretical aspects*. University of Pennsylvania Press, 1967. http://books.google.co.uk/books?hl=en&lr=&id=6rigtdo0u2UC&oi=fnd&pg=PA3&dq=Depression:+clinical,+experimental+and+theoretical+aspects.+Beck+1967&ots=T24fJ6IZ4J&sig=thZDfH7W5RLsXQthY_WD_oLjKj8 (accessed 13 Jan 2015).
58 Young JE, Klosko JS, Weishaar ME. *Schema therapy: a practitioner's guide*. Guilford Press, 2003. http://books.google.co.uk/books?hl=en&lr=&id=vScjGGgJEZgC&oi=fnd&pg=PA1&ots=hSWWcyJj_I&sig=a5AuhdJMCZYpL0AHeTeL17_jnwI (accessed 13 Jan 2015).
59 Maier SF, Watkins LR. Stressor controllability and learned helplessness: the roles of the dorsal raphe nucleus, serotonin, and corticotropin-releasing factor. *Neurosci Biobehav Rev* 2005;29:829–41.
60 Huys QJ, Dayan P. A Bayesian formulation of behavioral control. *Cognition* 2009;113:314–28.
61 Amat J, Baratta MV, Paul E, *et al*. Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nat Neurosci* 2005;8:365–71.
62 Killcross S, Coutureau E. Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb Cortex* 2003;13:400–8.
63 Huys QJ, Eshel N, O'Nions E, *et al*. Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol* 2012;8:e1002410.
64 Fales CL, Barch DM, Rundle MM, *et al*. Altered emotional interference processing in affective and cognitive-control brain circuitry in major depression. *Biol Psychiatry* 2008;63:377–84.
65 Joormann J, Gotlib IH. Emotion regulation in depression: relation to cognitive inhibition. *Cogn Emot* 2010;24:281–98.
66 Siegle GJ, Steinhauer SR, Thase ME, *et al*. Can't shake that feeling: event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. *Biol Psychiatry* 2002;51:693–707.
67 Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. *Perspect Psychol Sci* 2008;3:400–24.
68 Adams RA, Stephan KE, Brown HR, *et al*. The computational anatomy of psychosis. *Front Psychiatry* 2013;4:47.
69 Stephan KE, Baldeweg T, Friston KJ. Synaptic plasticity and dysconnection in schizophrenia. *Biol Psychiatry* 2006;59:929–39.
70 Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci (Regul Ed)* 2005;9:474–80.
71 Turetsky BI, Calkins ME, Light GA, *et al*. Neurophysiological endophenotypes of schizophrenia: the viability of selected candidate measures. *Schizophr Bull* 2007;33:69–94.
72 Anticevic A, Cole MW, Repovs G, *et al*. Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. *Cereb Cortex* 2014;24:3116–30.
73 Silverstein SM, Keane BP. Perceptual organization impairment in schizophrenia and associated brain mechanisms: review of research from 2005 to 2010. *Schizophr Bull* 2011;37:690–9.
74 Dima D, Roiser JP, Dietrich DE, *et al*. Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. *Neuroimage* 2009;46:1180–6.
75 Shergill SS, Samson G, Bays PM, *et al*. Evidence for sensory prediction deficits in schizophrenia. *Am J Psychiatry* 2005;162:2384–6.
76 Hong LE, Turano KA, O'Neill H, *et al*. Refining the predictive pursuit endophenotype in schizophrenia. *Biol Psychiatry* 2008;63:458–64.
77 Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* 2015;18:199–209.
78 Dickinson D, Ramsey ME, Gold JM. Overlooking the obvious: a meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Arch Gen Psychiatry* 2007;64:532–42.
79 Howes OD, Kapur S. The dopamine hypothesis of schizophrenia: version III—the final common pathway. *Schizophr Bull* 2009;35:549–62.
80 Demjaha A, Murray RM, McGuire PK, *et al*. Dopamine synthesis capacity in patients with treatment-resistant schizophrenia. *Am J Psychiatry* 2012;169:1203–10.
81 Goto Y, Grace AA. Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat Neurosci* 2005;8:805–12.
82 Dreyer JK, Herrik KF, Berg RW, *et al*. Influence of phasic and tonic dopamine release on receptor activation. *J Neurosci* 2010;30:14273–83.
83 Beaulieu JM, Gainetdinov RR. The physiology, signaling, and pharmacology of dopamine receptors. *Pharmacol Rev* 2011;63:182–217.
84 Smith Y, Bevan MD, Shink E, *et al*. Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience* 1998;86:353–87.
85 Ott T, Jacob SN, Nieder A. Dopamine receptors differentially enhance rule coding in primate prefrontal cortex neurons. *Neuron* 2014;84:1317–28.
86 Groman SM, Lee B, London ED, *et al*. Dorsal striatal D2-like receptor availability covaries with sensitivity to positive reinforcement during discrimination learning. *J Neurosci* 2011;31:7291–9.
87 Clatworthy PL, Lewis SJ, Brichard L, *et al*. Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory. *J Neurosci* 2009;29:4690–6.
88 den Ouden HE, Daw ND, Fernandez G, *et al*. Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* 2013;80:1090–100.
89 Leeson VC, Robbins TW, Matheson E, *et al*. Discrimination learning, reversal, and set-shifting in first-episode schizophrenia: stability over six years and specific associations with medication type and disorganization syndrome. *Biol Psychiatry* 2009;66:586–93.
90 Schlagenhauf F, Huys QJ, Deserno L, *et al*. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage* 2014;89:171–80.
91 Huys QJM, Tobler PN, Hasler G, *et al*. Chapter 3—the role of learning-related dopamine signals in addiction vulnerability. In: Diana M, Di Chiara G, Spano P, eds. *Progress in brain research*. Elsevier, 2014:31–77. http://www.sciencedirect.com/science/article/pii/B9780444634252000039 (accessed 28 Jan 2015).
92 Sesack SR, Carr DB. Selective prefrontal cortex inputs to dopamine cells: implications for schizophrenia. *Physiol Behav* 2002;77:513–17.
93 Fusar-Poli P, Howes OD, Allen P, *et al*. Abnormal prefrontal activation directly related to pre-synaptic striatal dopamine dysfunction in people at clinical high risk for psychosis. *Mol Psychiatry* 2011;16:67–75.
94 Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science* 1997;275:1593–9.
95 Lammel S, Ion DI, Roeper J, *et al*. Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron* 2011;70:855–62.
96 Friston K, Schwartenbeck P, Fitzgerald T, *et al*. The anatomy of choice: active inference and agency. *Front Hum Neurosci* 2013;7:598.
97 Steinberg EE, Keiflin R, Boivin JR, *et al*. A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* 2013;16:966–73.
98 Winton-Brown TT, Fusar-Poli P, Ungless MA, *et al*. Dopaminergic basis of salience dysregulation in psychosis. *Trends Neurosci* 2014;37:85–94.
99 Murray GK, Corlett PR, Clark L, *et al*. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol Psychiatry* 2007;13:267–76.
100 Corlett PR, Murray GK, Honey GD, *et al*. Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain* 2007;130:2387–400.
101 Murray GK, Corlett PR, Fletcher PC. The neural underpinnings of associative learning in health and psychosis: how can performance be preserved when brain responses are abnormal? *Schizophr Bull* 2010;36:465–71.
102 Averbeck BB, Evans S, Chouhan V, *et al*. Probabilistic learning and inference in schizophrenia. *Schizophr Res* 2011;127:115–22.
103 Deserno L, Huys QJ, Boehme R, *et al*. Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proc Natl Acad Sci USA* 2015;112:1595–600.
104 Collins AG, Brown JK, Gold JM, *et al*. Working memory contributions to reinforcement learning impairments in schizophrenia. *J Neurosci* 2014;34:13747–56.
105 Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology (Berl)* 2007;191:391–431.
106 Flagel SB, Clark JJ, Robinson TE, *et al*. A selective role for dopamine in stimulus-reward learning. *Nature* 2011;469:53–7.
107 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.
108 Roiser JP, Howes OD, Chaddock CA, *et al*. Neural and behavioral correlates of aberrant salience in individuals at risk for psychosis. *Schizophr Bull* 2013;39:1328–36.

109 Juckel G, Schlagenhauf F, Koslowski M, *et al*. Dysfunction of ventral striatal reward prediction in schizophrenia. *Neuroimage* 2006;29:409–16.
110 Hartmann MN, Hager OM, Reimann AV, *et al*. Apathy but not diminished expression in schizophrenia is associated with discounting of monetary rewards by physical effort. *Schizophr Bull* 2015;41:503–12.
111 Browning M, Behrens TE, Jocham G, *et al*. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci* 2015;18:590–6.
112 Lawson RP, Rees G, Friston KJ. An aberrant precision account of autism. *Front Hum Neurosci* 2014;8:302.
113 Hauser TU, Iannaccone R, Ball J, *et al*. Role of the medial prefrontal cortex in impaired decision making in juvenile attention-deficit/hyperactivity disorder. *JAMA Psychiatry* 2014;71:1165–73.
114 Edwards MJ, Adams RA, Brown H, *et al*. A Bayesian account of "hysteria". *Brain* 2012;135:3495–512.
115 Wells A. A cognitive model of generalized anxiety disorder. *Behav Modif* 1999;23:526–55.
116 Strauss GP, Gold JM. A new perspective on anhedonia in schizophrenia. *Am J Psychiatry* 2012;169:364–73.
117 Palminteri S, Lebreton M, Worbe Y, *et al*. Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. *Proc Natl Acad Sci USA* 2009;106:19179–84.
118 Bonnelle V, Veromann KR, Burnett Heyes S, *et al*. Characterization of reward and effort mechanisms in apathy. *J Physiol Paris* 2015;109:16–26.
119 Pareés I, Brown H, Nuruki A, *et al*. Loss of sensory attenuation in patients with functional (psychogenic) movement disorders. *Brain* 2014;137:2916–21.