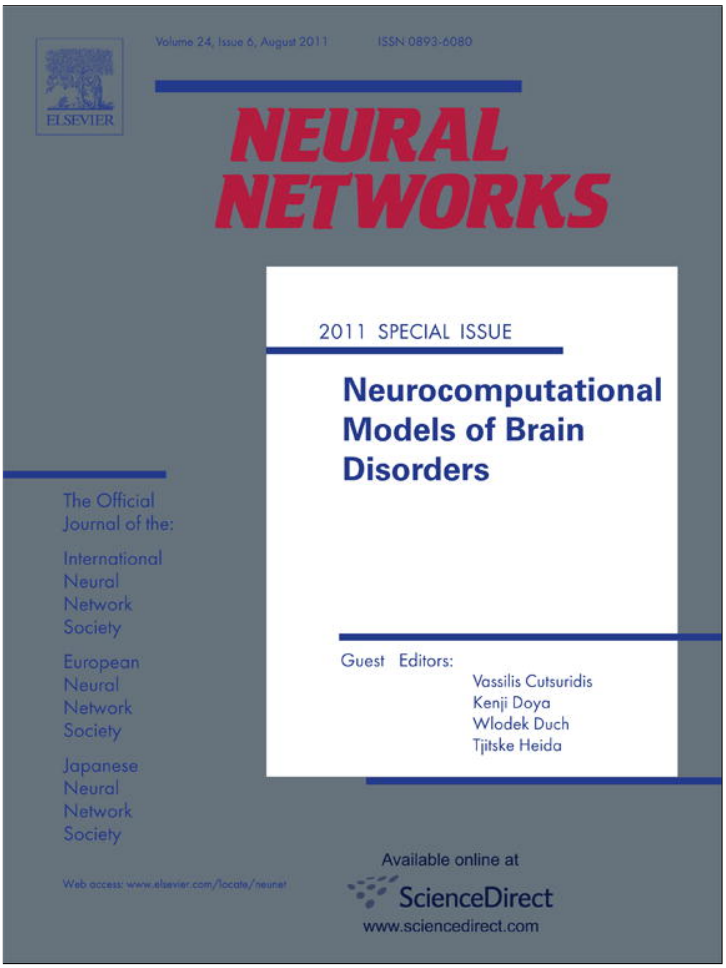
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Neural Networks 24 (2011) 544–551

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2011 Special Issue

Are computational models of any use to psychiatry?

Quentin J.M. Huys [a,*,1], Michael Moutoussis [b,1], Jonathan Williams [c,1]

[a] Wellcome Trust Centre for Neuroimaging, Gatsby Computational Neuroscience Unit and Medical School, UCL, United Kingdom
[b] School of Psychology, University of Manchester, United Kingdom
[c] Institute of Psychiatry, King's College, London, United Kingdom

ARTICLE INFO

Keywords:
Computational psychiatry
Computational neuroscience
Psychiatry
Computational models

ABSTRACT

Mathematically rigorous descriptions of key hypotheses and theories are becoming more common in neuroscience and are beginning to be applied to psychiatry. In this article two fictional characters, Dr. Strong and Mr. Micawber, debate the use of such computational models (CMs) in psychiatry. We present four fundamental challenges to the use of CMs in psychiatry: (a) the applicability of mathematical approaches to core concepts in psychiatry such as subjective experiences, conflict and suffering; (b) whether psychiatry is mature enough to allow informative modelling; (c) whether theoretical techniques are powerful enough to approach psychiatric problems; and (d) the issue of communicating clinical concepts to theoreticians and vice versa. We argue that CMs have yet to influence psychiatric practice, but that they help psychiatric research in two fundamental ways: (a) to build better theories integrating psychiatry with neuroscience; and (b) to enforce explicit, global and efficient testing of hypotheses through more powerful analytical methods. CMs allow the complexity of a hypothesis to be rigorously weighed against the complexity of the data. The paper concludes with a discussion of the path ahead. It points to stumbling blocks, like the poor communication between theoretical and medical communities. But it also identifies areas in which the contributions of CMs will likely be pivotal, like an understanding of social influences in psychiatry, and of the co-morbidity structure of psychiatric diseases.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past two decades, computational and theoretical approaches have blossomed in many fields related to the brain, ranging from cellular and network neuroscience to cognitive psychology and behavioural economics. More recently attempts have also been made to apply computational approaches to psychiatry. However, their usefulness for psychiatry is as yet uncertain. Here we debate both sides of the argument. We present a discussion between a sceptic and an enthusiast. In a final section, we will attempt to map out a useful way forward informed by these different arguments.

Two definitions are needed before we begin. First, we will use "psychiatry" for "mental health care" very generally. Second, the term "computational model" (CM) conventionally means any model on a computer. This overly broad category is meaningless because it encompasses any model that one might test using a statistical package on a computer. We will concentrate on a much narrower definition of CMs as models which express concepts important to brain function, and the relationships between these concepts, in mathematical form. That is, we will discuss whether attempting to phrase psychiatric concepts explicitly in the language of mathematics and statistics is likely to usefully advance our understanding of psychiatric problems. The kinds of concepts we have in mind are anhedonia, thought disorder, impulsivity, paranoia, etc. We admit a certain predilection for reinforcement learning models (Montague, Dayan, & Sejnowski, 1996; Sutton & Barto, 1998) and will use several examples from our own work. However, we do not intend our arguments to be specific to a particular form of model; rather, these models are highly diverse and depend on the particular problem at hand. They include neurally explicit (Ahmed, Graupner, & Gutkin, 2009) and abstract (McClelland, Rumelhart, & Hinton, 1986) connectionist models; information theoretic (Srinivasan, Laughlin, & Dubs, 1982), more explicitly statistical (Dayan, Hinton, Neal, & Zemel, 1995; Hinton, Dayan, Frey, & Neal, 1995; Hopfield, 1982) and optimal inference (Fiser, Berkes, Orbán, & Lengyel, 2010; Tenenbaum, Griffiths, & Kemp, 2006) models and game theoretic (Camerer, 2003; Fehr & Schmidt, 1999) models.

Wise old Dr. Strong (Dickens, 1850) will now put the case against CMs from the point of view of a psychiatrist. Our optimistic – or maybe unrealistic – friend Mr. Micawber will try to enthuse

* Corresponding author.
E-mail addresses: qhuys@cantab.net (Q.J.M. Huys), fzsemmo@gn.apc.org (M. Moutoussis), johwilliams@gmail.com (J. Williams).
URL: http://www.gatsby.ucl.ac.uk/~qhuys (Q.J.M. Huys).

[1] All authors contributed equally.

0893-6080/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2011.03.001

Author's personal copy

him about their cause. He is also a fan of reinforcement learning models.

Dr. Strong: In brief, the problems with CMs are as follows. First, CMs are fundamentally flawed in that they do not capture core clinical concerns. Second and third, the state of psychiatry and modelling are each too immature to be usefully applied to the other. Fourth, there are serious difficulties in the way information and ideas are transmitted between modellers and psychiatrists.

Mr. Micawber: These are important points, let us debate them one by one.

2. Are CMs clinically relevant?

Dr. Strong: First and foremost, CMs have failed to influence clinical practice.

Mr. Micawber: I would agree, Dr. Strong, that CMs have not influenced clinical practice to date; but neither have most advances in neurosciences. In fact, we believe that CMs will be instrumental in helping to bridge the gap between neurobiology and psychiatry because CMs are able to link levels of descriptions and make well-founded predictions at one level based on information at another level.

Dr. Strong: I disagree. The question is: are they clinically relevant, not will they be at some point in the future. All the models omit the very centre of psychiatry: subjective experiences. No one I have met believes that computers feel duty, personal bonds, or sexual titillation.

Mr. Micawber: Although intuitive, the notion that computational models cannot "feel" is neither important nor indeed falsifiable. To the extent that internal states correspond to linguistic or other behaviours (e.g. a questionnaire measure of mood), they can be investigated by CMs as well as by any other method. Furthermore, external aspects of internal percepts are of interest even if they make no reference to internal states. Behaviours are strongly affected by objective, external influences. People pay for psychiatric help partly because internal subjective experiences have external objective correlates: because they cannot work or look after their children, not just because they feel sad (Wing et al., 1998).

Dr. Strong: I'm afraid, Mr. Micawber, that this does not quite deal with the problem. What gets psychiatrists involved is pain, suffering and conflict. Almost no computational models take this into account. The rare exceptions that do have not translated into clinical practice, or even clinical research.

Mr. Micawber: I think, Dr. Strong, that if you look at modern models of depression or psychosis you will see that they increasingly investigate experiences central to pain and suffering. Take for instance anhedonia, helplessness and paranoid anxiety. Both anhedonia and helplessness have been highly instrumental in the development of both therapies (Beck, Rush, Shaw, & Emery, 1979) and animal models (Maier & Watkins, 2005; Willner, 1997). The concepts of anhedonia and helplessness, however, make partially overlapping predictions: failure to reap rewards may be due both to an indifference to them, or to an inability to see that their attainment is under the individual's control. These can be disentangled using CMs. For instance, the same mathematical formalism was able to account for animal data (Huys & Dayan, 2009) and human decision making (Huys, Vogelstein, & Dayan, 2009). This additionally allowed helplessness to be demonstrated in humans in a pure reward situation. Paranoid anxiety is another area where CM research has engaged directly with psychic suffering. The expectation of highly aversive outcomes is central to paranoia (Bentall et al., 2009) and very unpleasant for patients. Several recent CMs of threat perception and avoidance have elucidated the complex roles of highly aversive expectations (Bentall et al., 2009) in paranoia (Moutoussis, Bentall, Williams, & Dayan, 2008; Smith, Becker, & Kapur, 2005; though see also Schmajuk & Zanutto, 1997).

Dr. Strong: This, *Mr. Micawber*, is still far from the richness of human, clinical reality. You have described CMs that focus on specific symptoms as being themselves pathological. However, the range of experiences found in healthy people (especially children), is vast. For example, on careful clinical examination 4% of 12-year-olds have experienced auditory hallucinations in the past 6 months (Horwood et al., 2008). It is not the symptom itself that is pathological, but its frequency, severity, or simply its negative impact on life. Absence of this last feature, the impact on life, dramatically reduces the relevance of CMs to clinical practice.

Mr. Micawber: Psychiatry at present does observe that this is the case; but just like CMs, it is unable to provide any insight into why this might be. Understanding the mechanisms that sustain psychiatric symptoms is more likely a step towards understanding their impact than the sole description of this fact.

Dr. Strong: A final point is that CMs need to address treatments, both pharmacological and learning based, rather than just focus on fundamental mechanisms. There is strong evidence for the effectiveness of these treatments, and if CMs are to have clinical utility, they need to contribute to our understanding of them.

Mr. Micawber: It is certainly true that CMs have, as yet, not made much inroad on talking therapies. We will, however, return to the issue of pharmacological therapies below.

3. Is psychiatry advanced enough for CMs?

Dr. Strong: It is far too early to mathematise psychiatry. Psychiatry is not advanced enough for rigorous CMs in at least two ways: current taxonomy is unsatisfactory and fundamental concepts are in flux. Categories in psychiatry (including those in DSM (American Psychiatric Association, 1994) and ICD (World Health Organization, 1990)) are at best heterogeneous, overlapping groups. Most modellers do not seem to be aware of the unfortunate state of academic psychiatry, which can currently be described as the inexorable, painfully slow, deconstruction of DSM categories.

Mr. Micawber: I think that you are going too far. Valid descriptions of the brain and its pathologies exist at multiple, independent levels (Clark, 2001; Kendler & Parnas, 2008; Marr, 1982). Many different mechanisms can produce the same function (as in convergent evolution); conversely, the same mechanism can produce many functions (as can computers, or brains). There is no necessary one-to-one mapping from a low-level mechanism to a high-level function, or vice versa. It would be a fundamental error to believe that data-driven concepts derived from epidemiological analyses are not valid just because there is no well-defined neurobiology corresponding to them; it is just as mistaken to believe that because a descriptive concept is statistically valid, there must be a corresponding neurobiology. The fact that DSM categories have not yet been convincingly matched to neurobiology is no more than a reflection of this.

Dr. Strong: Modellers are still attracted to constructs that clinician-researchers are becoming disenchanted with. Examples of such constructs include autism (Happe, Ronald, & Plomin, 2006), ADHD (Williams, 2010), "the schizophrenias" (Murray et al., 2005; Sanislow & Carson, 2001) and mood disorders (with DSM V proposing a new, partially merged, version of the depression and anxiety).

Mr. Micawber: You may be arguing against yourself, Dr. Strong. CMs that capture the entirety of these controversial syndromes are very far and few between. Which CM of depression, for instance, includes anhedonia and also problems with sleep and appetite? CMs are, by their very nature, focussed on key functions and their consequences, i.e., key dysfunctions. It may be tempting to treat a key dysfunction as representing the essence of a diagnostic

Author's personal copy

category. The status of diagnostic categories, however, is highly problematic; less so than the status of key dysfunctions (Bentall, 2003). As such, far from the mismatch with diagnostic categories being indicator that the state of psychiatry precludes insights from modelling, it means that CMs have a greater relevance to mental health problems. CM researchers must take seriously the proposition that their (dys)function based approach is not a poor relation to the diagnostic one but may in fact help advance the latter!

Dr. Strong: I agree that this approach is attractive, but I'm convinced that it only applies to a tiny subset of biological processes in psychiatry. Though there may be a few exceptions for long-evolved organising principles (such as monoamines or simple drives; (Gray, 1991; Kapur, 2003; Panksepp, 1998; Williams, 2006; Wise, Berger, & Stein, 1972)), most problems in psychiatry are of a level of complexity that far exceeds that of simple behaviours and precludes any useful application of simple CMs. Let us not forget that we (and CMs) are still unable to describe all but the simplest aspects even of extremely well described, straightforward, behaviours such as operant conditioning.

Mr. Micawber: Fortunately, every single one of these statements is overly simplistic to the point of being incorrect. Long-evolved organising principles of the brain are of profound importance for psychiatry; and emphasizing monoamine function is the rule, not the exception, in biological psychiatry. Furthermore, let us not forget that the majority of treatments depend on monoamines. Arguing that understanding these issues is peripheral to psychiatry strikes me as frivolous. Let me just mention three examples:

Selective serotonin reuptake inhibitors (SSRIs) are widely used as antidepressants and anxiolytics, but reconciling their psychiatric effects with basic psychopharmacology has been problematic. Unlike their positive effects on mood in humans, SSRIs can enhance the aversive effects of punishments in animal models (Cools, Roberts, & Robbins, 2008; Dayan & Huys, 2009; Soubrié, 1986). We found that a simple CM was able to reconcile these two findings by linking aversive predictions to behavioural inhibition in an adaptive manner that had overall positive consequences (Dayan & Huys, 2008).

Paranoia and antipsychotic drugs: Next consider dopamine (DA), whose involvement in psychosis is very well established (Kegeles et al., 2008; Laruelle, 2008). Dopamine-sensitive animal models, such as the 'conditioned avoidance response' (Courvoisier, 1956), have long been used to assess 'antipsychotic' action. This, however, was done without any knowledge of how DA blocking drugs altered information processing in the brain. The advent of CMs based on Prediction-Error learning theory (Schultz, Dayan, & Montague, 1997; Sutton & Barto, 1998) first helped link DA blockers to brain information processing (Smith, Li, Becker, & Kapur, 2007). CMs then helped understand that DA may signal prediction errors about actions leading to both worse-than-expected and better-than-expected outcomes. This holds even if DA, a modulator hitherto related to rewards, has little role in signalling worse-than-expected situations (states). Modelling related the psychological mode of action of dopamine blocking drugs to dampening (1) the vigour of the avoidance response and (2) the prediction-error signals that drive action learning (Moutoussis et al., 2008).

Impulsivity: Dopamine function is also related to impulsivity, which in turn is central to many psychiatric conditions from attention-deficit hyperactivity disorder (ADHD) to borderline personality and substance abuse disorders. A modelling analysis of the behaviour of the spontaneously hypertensive rat (the most prominent animal model of ADHD) clarified how stimulants, by impacting on phasic dopamine, can impact on "impulsive" measures (Williams & Dayan, 2005). Further work has since clarified the role of tonic dopamine signals to motivation and aspects of impulsivity (Niv, Daw, Joel, & Dayan, 2007). All the while these CMs have emphasised quite how many different aspects of learning can influence impulsivity (Williams & Dayan, 2005). The concept of dopamine appetite, which emerged from this CM account of previous findings in humans (Williams & Taylor, 2004), has since received experimental verification in rats (Williams, Sagvolden, Taylor, & Sagvolden, 2009).

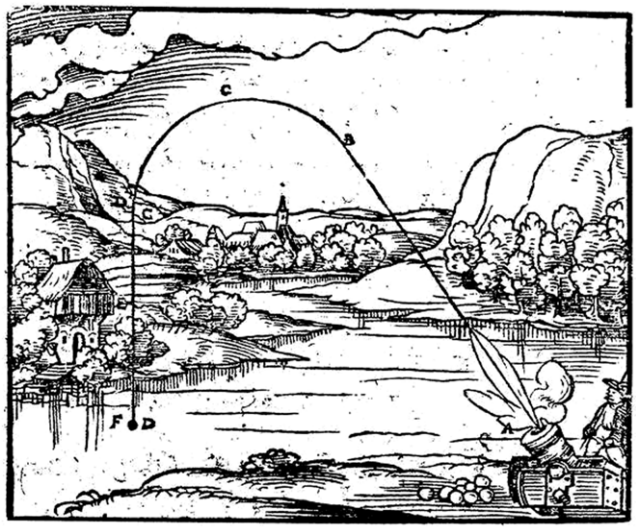


Fig. 1. The renaissance view of projectile motion was conceptually complex. It divided the trajectory into several parts (two straight and one curved!) and included an account of the interaction of the projectile with the atmosphere. Though intuitively satisfactory, it was computationally intractable and ultimately misleading. Image downloaded from http://tinyurl.com/34p9up6.

Dr. Strong: Returning to the overall point: psychiatry at present is still dominated by definitions, data collection and epidemiology rather than testable, risky, hypotheses. An example of what should happen more often are the predictions by Sonuga-Barke that children with ADHD would respond in a particular way during a matching familiar figures task. The predictions were sufficiently explicit at a mechanistic level that they could be tested by implementing them in a CM. In this case, the predictions were disproved (according to one view; Williams & Taylor, 2004) to the great credit of the original proponent who was brave enough not to make yet another risk-free "so what" hypothesis.

Mr. Micawber: Psychiatry is indeed a complicated, fragmented field teeming with partial explanations. It is precisely because of this that powerful principles are needed to unify the fragments of theory and increase explanatory power. At present psychiatric theories resemble the overly complicated, qualitative Aristotelian theories of projectile motion of the late renaissance (Fig. 1). The history of the study of gravity is informative. The breakthroughs of Galileo's generation did not just come through accumulation of knowledge. In modern terms, they came through the iterative development of a mathematical core theory and of experiments based on this core theory to arrive at important principles. Specifically, they (i) considered limiting cases where key approximations could be used; (ii) focussed experiments at these mathematically tractable cases; and (iii) rigorously related the mathematical descriptions to putative general principles. This led to the discovery of powerful general principles e.g. that free-fall motion is the same for all objects. Today, these principles are so obvious to us that it takes a lot of effort to persuade oneself that the Aristotelians weren't either electively blind or stupid. Deep insights about key principles make profoundly puzzling and highly complex findings (as they are prevalent in psychiatry today) look expected and commonsensical.

In mental health, our best guess is that the key principles involve information processing such as probabilistic inference and optimal decision making. The brain's raison d'être is to process

Author's personal copy

information in an uncertain world; and psychiatric phenomena do have an essential core of deviant information processing and inappropriate certainties: perceptions "without a stimulus", propositions "contrary to evidence", cognitive inferences that are "biased", etc. Information processing is a computational concept. Its many facets can be captured accurately and thoroughly by a variety of powerful, specific CMs (see Dayan & Abbott, 2001, for an overview); and these can be applied directly to psychiatric problems (e.g. Chater & Oaksford, 2008; Huys, 2007; Maia & Frank, 2011; Williams & Dayan, 2005). Crucially, these approaches link psychopathology to normal psychology and to basic neuroscience, for instance investigating how changes in GABA or NMDA signalling, supported by work in genetics and animal models, may explain the perceptual features of schizophrenia (Loh, Rolls, & Deco, 2007; Migliore, Blasi, Tegolo, & Migliore, 2011). Computational models, more than any other approach, allow us to relate findings to general principles that tap the core of the brain's raison d'être, which is to compute and process information, rather than say produce a stream of internal experiences.

Prediction-Error (PE) theory is a specific example of such an approach, which gave rise to several of the key insights regarding the monoamines mentioned above. It originated from the recognition that phasic activity of monkey mesolimbic DA neurons appears to report a signal much like one previously described in formal models of reinforcement learning (Rescorla & Wagner, 1972; Schultz, Apicella, & Ljungberg, 1993; Schultz et al., 1997; Sutton & Barto, 1998). This has since been replicated in electrophysiological and imaging studies in humans (D'Ardenne, McClure, Nystrom, & Cohen, 2008; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; Zaghloul et al., 2009) following unexpected reinforcements. Phasic DA activations conform in important details with the theoretical PE account (Waelti, Dickinson, & Schultz, 2001), being proportional to the difference between what was expected and what was found (Bayer & Glimcher, 2005)—hence the term PE. Precisely such signals can be used to learn expectations from experience (Montague et al., 1996), and also to learn new behaviours (Montague et al., 2004). Moreover in recent years the concept of PE has achieved a much wider significance as it has been realised that some brain signals are likely to be dopamine independent (e.g. Boureau & Dayan, 2011).

This account, although far from perfect, is an explicit and testable model of learning that bridges levels of description, linking high-level choice behaviour to detailed underlying neurobiological mechanisms. Its simplicity belies its computational depth and it has led to an explosion of research ideas linking psychiatry to normal neuroscience and a refinement of functional imaging methods. Finally, it has already generated specific predictions that have been experimentally tested (Redish, 2004, Panlilio et al., 2007; see also Maia & Frank, 2011).

4. Are CMs mature enough to face the complexities of psychiatry?

Dr. Strong: The fact remains that there still are no overarching computational theories that provide a convincing framework in which to approach human experience or behaviour. Existing computational theories are piecemeal, unrelated and do not come together in a global view.

I previously criticised psychiatrists for not making sufficiently explicit and risky predictions; a very similar criticism applies to CMs. The essence of science is to make surprising predictions, i.e., that are not expected based on prior knowledge. CMs fall down in both aspects of these requirements. First they are not surprising, second they do not make real predictions. An example of lack of prediction is that many CMs "predict" results that were already published.

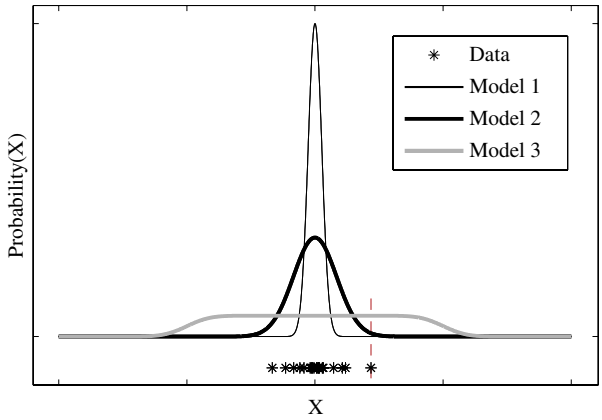


Fig. 2. Model complexity and evidence. See text for details.

Mr. Micawber: Risky predictions are not the *sine qua non* of research: medicine at large has greatly benefited from testing common assumptions. In a key paper, Echt et al. (1991) tested the commonly held belief that ventricular extrasystoles after myocardial infarction should be controlled, finding instead that treatment increased mortality. In terms of psychiatry, it has long been known that patients with schizophrenia are impaired on reversal learning tasks (Elliott, McKenna, Robbins, & Sahakian, 1995; Waltz & Gold, 2007). Original accounts emphasised the importance of the negative feedback at the reversal stage. Surprisingly and counter-intuitively, detailed modelling showed that subjects are mainly guided by rewards and that, in fact, it is learning from rewards that differentiates schizophrenia patients from controls (unpublished data, F. Schlagenhauf, Q. Huys and A. Heinz).

The traditional approach may in fact mislead researchers into making claims that are too risky by focussing only on one aspect of the data. For instance, Elliott, Sahakian, Herrod, Robbins, and Paykel (1997) elegantly showed that depressed subjects made many errors on trials following negative feedback. They interpreted this as a non-specific impairment of response to negative feedback. However, any such impairment would have a number of global effects both on other aspects of the task, and on other tasks. Building models can enforce consideration of the global consequences of any hypothesis, thereby facilitating comprehensive testing of hypotheses.

Dr. Strong: Most CMs produce no surprises because they are overly powerful: almost any result can be accounted for by "learning" mechanisms (Tripp & Wickens, 2008; Williams, 2008). In models that specify links between environment and psychopathology, the links are largely under-constrained and often so flexible that they can trivially account for any desired finding. The other side of this coin is that modellers also lighten their plight by neglecting aspects of data: rather than accounting for the entire distribution, they have a tendency to create a computational "type specimen" and then to match this to published group means, neglecting any population spread.

Mr. Micawber: In brief, Dr. Strong, you are talking about parsimony. It is indeed important to balance the power of a model against the complexity required of it by the data. Formal methods of model comparison (Kass & Raftery, 1995) can efficiently handle this, are already standard procedure in imaging (Stephan, Penny, Daunizeau, Moran, & Friston, 2009) and being applied to behavioural studies (Daw, 2011; Huys et al., in press, 2009). Consider the dataset displayed as stars (*) in Fig. 2. The question we would like to answer is which of the three models best accounts

Author's personal copy

for this data. The three distributions show what type of data each model can generate. Model 1 makes quite specific predictions. It is thus a good, easily falsifiable model. The point indicated by a dotted line for instance is very unlikely under Model 1. On the other hand, Model 3 stands for a more complex model with more parameters. The increase in complexity makes it more powerful in that it can generate a broad variety of data. Here, it seems to account very well for all the data, in that all the points that are observed have equal and high probability under that model. Unfortunately, though, Model 3 also makes many predictions which are not observed; as a direct consequence of spreading its predictions too widely, it assigns less likelihood to the dense concentration of data in the centre than the other two models. Model 2 presents the best trade-off. Quantities which normatively compare models by trading off their complexity against the complexity needed to account for the data can be computed explicitly for CMs, allowing for a thorough evaluation. For a more thorough, yet concise, exposition of model comparisons, see MacKay (2003, Chapter 28). Furthermore, such an approach clearly take a model's ability to account for the distribution of observed data very seriously.

Dr. Strong: Taking a broader view, we can ask what CMs could possibly aim to achieve. When simple counter-examples suffice, CMs are tools that are useful to disprove categorical statements (as in the Sonuga-Barke example above). However, innumerable small assumptions flow into the building of any CM; some explicitly stated, others not. Subtle aspects of the models are invariably carefully tailored to the task in a manner, which is hard to test or assess objectively. For instance, it is well known that the performance of reinforcement learning models depends critically on the formulation of the state-space; an issue seldom if ever made explicit.

Mr. Micawber: On the contrary, it is non-mathematical theories that best hide their – often complex – assumptions! Building models allow for thorough assessment of parsimony. It is difficult to implement a computational model without making explicit many assumptions that traditional psychiatric research overlooks. When these are taken into account, important, unexpected predictions often emerge. Consider the following examples:

First, in probabilistic reinforcement tasks, sometimes feedback will be correct (i.e., informative of the true correct response), and at other times not. Arguing that there is a difference in the response to "erroneous" and "informative" feedback (e.g. Cools, Clark, Owen, & Robbins, 2002) in reversal learning implies that subjects know which is which. This in turn implies a much more complex inferential model (a hidden state model), which makes many additional predictions that may or may not be supported.

Second, it has long been known that deluded patients utilise only a small amount of sequentially presented information to reach a decision, a phenomenon labelled "jumping to conclusions" (Fine, Gardner, Craigie, & Gold, 2007). A predominant explanation is that paranoid patients somehow associate greater costs with sampling more information (Bentall, 2003). Using a CM, however, it was found that the detailed pattern of patient's responses is not consistent with this explanation. Cognitive noise, known about but overlooked in traditional research, is higher in deluded patients and actually accounts for the data (Moutoussis, Bentall, El-Deredy, & Dayan, in press). This study also carefully took into account within-population variability and included it in model testing (see Fig. 2).

In both these examples, detailed implementation of the assumed inference processes revealed that the commonsense account was incorrect; and it revealed novel processes that do account for the data better.

5. Modellers do not communicate well to psychiatrists

Dr. Strong: Papers on computational models are difficult to read and even more difficult to replicate (even though this should in principle be a simple matter of running a program). Possibly because of this, CMs tend to remain isolated from prior findings, and even from prior CMs. Rather than converging towards more holistic models of human psychology, the field fragments into ever more elaborate CMs of smaller topics. This is compounded by a further key problem: that mathematics allows for more complexity, and thereby produces models that are as hard to understand as the original problem. In fact, some models are so complicated that hardly anyone understands them, for instance Grossberg's impressively detailed model of neurobiological dysfunction underlying autism (Grossberg & Seidman, 2006).

Mr. Micawber: Reverse engineering is one possible use of models where the aim is to replicate the original system as far as possible. It is true that such models may ultimately be very complex and difficult to understand. However, such models represent the synthesis of many smaller, reductionist models you criticised earlier. By allowing for complex, multifaceted, explanations, CM techniques expand traditional models of explanation in psychiatry and as such should be a welcome addition to the toolbox.

Dr. Strong: Many CMs rely on intuitions germane to mathematicians, but do so at the expense of being intuitive to psychiatrists. The notion of "Bayesian" is an example. Although the basic theorem is simple and elegant, the deeper issues involving the computational complexities of performing say Bayesian model comparison make it at best an extremely cumbersome tool.

Mr. Micawber: You can't expect CMs to be free of maths or mathematical concepts. That would be like asking a paper on schizophrenia to be free of patients. If psychiatrists were familiar with mathematics, CMs would indeed play a more important role in psychiatric research. On the other hand this is far from being a precondition for CMs to flourish in psychiatry. In important areas (such as imaging in psychiatry research; Knutson, Bhanji, Cooney, Atlas, & Gotlib, 2008; Menon et al., 2007) CMs are already routinely used; they have become Kuhnian "normal science" (Notturno, 1984). Psychiatrists participate in such research and enjoy its fruits just as in the case of many other other tools from CPUs to MRI scanners.

Dr. Strong: Well, but in terms of the kinds of CMs we agreed to talk about (i.e., CMs that use maths to model key concepts), not even computational labs work with one another's models. How could non-computational labs or clinicians? Some of these communication difficulties may of course be overcome by clear writing and with the help of collaborators from other disciplines. But at present, psychiatric readers cannot be sure whether the obscurity of CM results is due to the complexity of the ideas, to an alternative but reasonable view of humans, or just down to error or irrelevance.

Mr. Micawber: For once I agree with you. We can conclude our debate on this note: that communication and collaboration in CM research needs to be thought about more.

6. Discussion

Computational models are powerful tools. Their affinity for fundamental issues in brain function makes them particularly strong candidates for research on neurobiological and psychological bases of psychiatric disorders. This link should be useful for integrating psychiatry with neuroscience, and for giving psychiatry more of an overarching framework. Their mathematical and statistical form lends them to hypothesis testing at an exacting level. This allows for thorough comparisons in both qualitative and quantitative terms, the former by producing qualitative predictions through

Author's personal copy

simulations, the latter by using advanced methods to compare the ability of one theory to explain the data better than alternative theories (Daw, 2011; Huys et al., in press; Williams & Taylor, 2004). This mathematical and statistical nature of CMs makes it quite practicable to assess their predictive strength (MacKay, 2003). However, computational models are not a panacea for psychiatry's many ills and they share the limitations of many other methodologies. First, CMs will not find a mapping between levels of (biopsychosocial) organisation if such a mapping does not exist. CMs can only help us discover the neurobiology of a DSM category if it exists. Second, although CMs can be used to extract more from data, limitations of measurements will persist. If our measurement of a variable is poor, this will hinder CMs just as much as it hinders any other research method. Third, computational models can at times be brought to bridge levels of description and bring many constraints to bear on a problem; but this does not remotely approach the kind of multifaceted view that a psychiatrist takes of a patient.

In addition, computational models do suffer from limitations specific to them. Many concepts can be hard to put in mathematically precise terms. At times, this may point to an immaturity of the concept, but at times it is also indicative of a limitation of the mathematical tools available. For example, although idiosyncrasies in language can be described with highly complex language models from machine learning (e.g. (Blei & Lafferty, 2009)), we are still very far from being able to integrate such powerful, yet only descriptive, models with CMs in neuroscience, let alone psychiatry. Next, although performing information processing better than any computer, brains are made of neurons. CMs relevant to psychiatry need to incorporate neurobiological constraints. As such, they suffer from the limitations arising from the state of neuroscience, and indeed by the state of (neuro)biological psychiatry.

Finally, many problems which appear natural to us are computationally very complex. The fact that computers only recently beat humans at chess emphasises the immature state of our understanding of computational problems. It may be that we have to await further advances in purely computational research before mental health practitioners can hope to see any consequences in their clinic. Of course, all these issues compound each other: for instance, the computational problems faced in testing and implementing a model of addiction would be rendered even less tractable by adding constraints from research into the dozens of metabolic pathways involved (Volkow & Li, 2004).

6.1. Future directions

Several important issues, both obstacles and opportunities, need to be addressed for computational psychiatry to flourish.

The key obstacle that has to be overcome is the difficulty in communication between theoreticians and empirical researchers. In psychiatry, empirical researchers naturally prefer to work on their own theories and, not heeding the advice of Freeman (1992), only rarely see the relevance of modellers. Theoretician modellers cannot expect psychiatrists to study formulae; they need to distil the qualitative significance of their work and to communicate it powerfully. They need to claim a place in the actual design of experiments, in the applications for research grants to carry out these experiments, and the CM based analysis of data. Computational psychiatrists must address questions of consequence to clinical and experimental practice, as computational cardiologists are doing (Rudy et al., 2008).

Second, in parallel to new data being gathered, old data can be exploited. A huge amount of data exists already. Many neuropsychological tasks, for instance, are straightforward to model. The tasks may indeed have neurobiological correlates, and models can help us to sharpen any inferences (Moutoussis, Orrell, & Morris, 2004). This is particularly true because models can be built that span several tasks. Making extant data more easily accessible would attract talented modellers to the task and improve the quality and usefulness of models very rapidly.

Models will further help integrate the levels of psychiatric description. It will be critically important to acknowledge the partial independence of the levels of description: few differences between bicycles are captured by whether they are made of steel or titanium. There likely is no one-to-one mapping between neurobiological mechanisms and most DSM categories—and we do not need to reify the latter. Indeed, focussing on specific functions that are computationally tractable but also important for psychiatry, helps integrating the insights of different disciplines. This transdiagnostic approach is fully in tune with experimental psychology and genetics (Bentall, 2003; Craddock, O'Donovan, & Owen, 2006).

The interaction between functional brain systems will be crucial to any understanding of psychiatric problems. A first stage is to demonstrate how multiple functional systems could interact. In mood disorders, we have suggested that the complexity of decision making is central (Huys, 2007). Because high-level decision making systems must rely on low-level building blocks, any disruption at a low level will have profound repercussions for the higher-level systems. Interactions between different systems (and indeed between systems contributing to different endophenotypes) arising from a sharing of such basic building blocks are likely to go some way towards explaining both symptom and treatment overlap between syndromes. The interaction between the decision systems is a complex computational problem constrained by nature's possibly haphazard choices. As such, it is a prime object for theoreticians and experimentalists to collaborate on.

Social interactions are very important in psychiatry, as they are in economics. Behavioural economists have elucidated many issues of great psychiatric interest, such as cooperation, fairness, altruism and selfishness (deQuervain et al., 2004; Fehr & Schmidt, 1999). As many of the social probes, or "games", used by behavioural economists have been subject to rigorous mathematical analysis (Camerer, 2003; Yoshida, Dolan, & Friston, 2008), this approach can be naturally extended to illuminate the basis of social dysfunction in psychiatric disorders (King-Casas et al., 2008; Meyer-Lindenberg, Mervis, & Berman, 2006). In the case of personality disorder the computational framework is still being developed (Ray, King-Casas, Montague, & Dayan, 2009), but the approach has already helped to quantify the dysfunction in other people's mind-reading that autistic adults display (Koshelev, Lohrenz, Vannucci, & Montague, 2010; Yoshida et al., 2010).

6.2. Conclusion

Strictly speaking, all models are wrong; but some are still very useful (Box & Draper, 1986). At their worst, computational models incomprehensibly regurgitate known facts. Sometimes they are a useful statistical tool that allows complex hypotheses to be rigorously tested. At their best, they allow apparently discordant facts to come together in a harmonious novel framework, inspiring clinicians and scientists alike.

References

Ahmed, S. H., Graupner, M., & Gutkin, B. (2009). Computational approaches to the neurobiology of drug addiction. *Pharmacopsychiatry*, *42*(Suppl. 1), S144–S152.

American Psychiatric Association, (1994). *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association Press.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *The Guilford clinical psychology and psychotherapy series*, *Cognitive therapy of depression* (1st ed.) New York: Guilford Press.

Author's personal copy

Bentall, R. P. (2003). *Madness explained*. London, UK: Penguin Books.

Bentall, R. P., Rowse, G., Shryane, N., Kinderman, P., Howard, R., Blackwood, N., et al. (2009). The cognitive and affective structure of paranoid delusions: a transdiagnostic investigation of patients with schizophrenia spectrum disorders and depression. *Archives of General Psychiatry*, *66*(3), 236–47.

Blei, D., & Lafferty, J. (2009). Topic models. In *Text mining: theory and applications*. Taylor and Francis.

Boureau, Y. L., & Dayan, P. (2011). Opponency revisited: competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology*, *36*(1), 74–97.

Box, G., & Draper, N. (1986). *Wiley series in probability and statistics, Empirical model-building and response surface*. Wiley-Blackwell.

Camerer, C. F. (2003). *Behavioral game theory: experiments in strategic interaction*. Princeton University Press.

Chater, N., & Oaksford, M. (Eds.). (2008). *The probabilistic mind: prospects for a Bayesian cognitive science*. Oxford: Oxford University Press.

Clark, A. (2001). *Mindware: an introduction to the philosophy of cognitive science*. New York: Oxford University Press.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *22*(11), 4563–4567. URL: http://dx.doi.org/20026435.

Cools, R., Roberts, A. C., & Robbins, T. W. (2008). Serotoninergic regulation of emotional and behavioural control processes. *Trends in Cognitive Sciences*, *12*(1), 31–40.

Courvoisier, S. (1956). Pharmacodynamic basis for the use of chlorpromazine in psychiatry. *Journal of Clinical and Experimental Psychopathology*, *17*(1), 25–37.

Craddock, N., O'Donovan, M. C., & Owen, M. J. (2006). Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. *Schizophrenia Bulletin*, *32*(1), 9–16.

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, *319*(5867), 1264–1267.

Daw, N. (2011). Trial-by-trial data analysis using computational models. In E. Phelps, T. Robbins, & M. Delgrado (Eds.), *Attention and performance XXIII*. Oxford: Oxford University Press.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience. Computational and mathematical modeling of neural systems*. MIT Press.

Dayan, P., Hinton, G., Neal, R., & Zemel, R. (1995). The Helmholtz machine. *Neural Computation*, *7*(5), 889–904.

Dayan, P., & Huys, Q. J. M. (2008). Serotonin, inhibition, and negative mood. *PLoS Computational Biology*, *4*(2), e4.

Dayan, P., & Huys, Q. J. M. (2009). Serotonin in affective control. *Annual Review of Neuroscience*, *32*, 95–126.

deQuervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science*, *305*(5688), 1254–1258.

Dickens, C. (1850). David copperfield. *Internet Archive*,.

Echt, D. S., Liebson, P. R., Mitchell, L. B., Peters, R. W., Obias-Manno, D., Barker, A. H., et al. (1991). Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The cardiac arrhythmia suppression trial. *New England Journal of Medicine*, *324*(12), 781–788.

Elliott, R., McKenna, P. J., Robbins, T. W., & Sahakian, B. J. (1995). Neuropsychological evidence for frontostriatal dysfunction in schizophrenia. *Psychological Medicine*, *25*(3), 619–630.

Elliott, R., Sahakian, B. J., Herrod, J. J., Robbins, T. W., & Paykel, E. S. (1997). Abnormal response to negative feedback in unipolar depression: evidence for a diagnosis-specific impairment. *Journal of Neurology, Neurosurgery, and Psychiatry*, *63*, 74–82.

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation*. *Quarterly Journal of Economics*, *114*(3), 817–868.

Fine, C., Gardner, M., Craigie, J., & Gold, I. (2007). Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cognitive Neuropsychiatry*, *12*(1), 46–77.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119–130.

Freeman, W. (1992). Chaos in psychiatry. *Biological Psychiatry*, *31*(11), 1079–1081.

Gray, J. A. (1991). *Problems in the behavioural sciences*: *Vol. 5. The psychology of fear and stress* (2nd ed.) Cambridge, UK: Cambridge University Press.

Grossberg, S., & Seidman, D. (2006). Neural dynamics of autistic behaviors: cognitive, emotional, and timing substrates. *Psychological Review*, *113*(3), 483–525.

Happe, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, *9*(10), 1218–1220.

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The wake-sleep algorithm for unsupervised neural networks. *Science*, *268*(5214), 1158–1161.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8), 2554.

Horwood, J., Salvi, G., Thomas, K., Duffy, L., Gunnell, D., Hollis, C., et al. (2008). Iq and non-clinical psychotic symptoms in 12-year-olds: results from the alspac birth cohort. *British Journal of Psychiatry*, *193*(3), 185–191. URL: http://dx.doi.org/10.1192/bjp.bp.108.051904.

Huys, Q. J. M. (2007). Reinforcers and control. towards a computational ætiology of depression. *Ph.D. thesis*. Gatsby Computational Neuroscience Unit. UCL. University of London. URL: http://www.gatsby.ucl.ac.uk/~qhuys/pub.html.

Huys, Q. J. M., Cools, R., Friedel, M. G. E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Computational Biology* (in press).

Huys, Q. J. M., & Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition*, *113*(3), 314–328.

Huys, Q. J. M., Vogelstein, J., & Dayan, P. (2009). Psychiatry: insights into depression through normative decision-making models. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems*: *Vol. 21* (pp. 729–736). MIT Press.

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry*, *160*(1), 13–23.

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430).

Kegeles, L. S., Slifstein, M., Frankle, W. G., Xu, X., Hackett, E., Bae, S.-A., et al. (2008). Dose-occupancy study of striatal and extrastriatal dopamine d2 receptors by aripiprazole in schizophrenia with pet and [18f]fallypride. *Neuropsychopharmacology*, *33*(13), 3111–3125. URL:http://dx.doi.org/10.1038/npp.2008.33.

Kendler, K., & Parnas, J. (Eds.). (2008). *Philosophical issues in psychiatry: explanation, phenomenology, and nosology*. Johns Hopkins Univ. Pr.

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science*, *321*(5890), 806–810.

Knutson, B., Bhanji, J. P., Cooney, R. E., Atlas, L. Y., & Gotlib, I. H. (2008). Neural responses to monetary incentives in major depression. *Biological Psychiatry*, *63*(7), 686–692. URL: http://dx.doi.org/10.1016/j.biopsych.2007.07.023.

Koshelev, M., Lohrenz, T., Vannucci, M., & Montague, P. R. (2010). Biosensor approach to psycopathology classification. *PLoS Computational Biology*, *6*(10), e1000966. URL: http://dx.doi.org/10.1371/journal.pcbi.1000966.

Laruelle, M. (2008). Dopamine and persecutory delusions. In D. Freeman, R. Bentall, & P. Garety (Eds.), *Persecutory delusions: assessment, theory and treatment* (pp. 239–266). Oxford: Oxford University Press.

Loh, M., Rolls, E. T., & Deco, G. (2007). A dynamical systems hypothesis of schizophrenia. *PLoS Computational Biology*, *3*(11), e228. URL: http://dx.doi.org/10.1371/journal.pcbi.0030228.

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge University Press.

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*(2), 154–162. URL: http://dx.doi.org/10.1038/nn.2723.

Maier, S. F., & Watkins, L. R. (2005). Stressor controllability and learned helplessness: the roles of the dorsal raphe nucleus, serotonin, and corticotropin-releasing factor. *Neuroscience & Biobehavioral Reviews*, *29*(4–5), 829–841.

Marr, D. (1982). *Vision*. New York, NY, USA: Freeman.

McClelland, J., Rumelhart, D., & Hinton, G. E. (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press.

Menon, M., Jensen, J., Vitcu, I., Graff-Guerrero, A., Crawley, A., Smith, M. A., et al. (2007). Temporal difference modeling of the blood–oxygen level dependent response during aversive conditioning in humans: effects of dopaminergic modulation. *Biological Psychiatry*, *62*, 765–772.

Meyer-Lindenberg, A., Mervis, C. B., & Berman, K. F. (2006). Neural mechanisms in Williams syndrome: a unique window to genetic influences on cognition and behaviour. *Nature Reviews Neuroscience*, *7*(5), 380–393.

Migliore, M., Blasi, I. D., Tegolo, D., & Migliore, R. (2011). A modeling study suggesting how a reduction in the context-dependent input on ca1 pyramidal neurons could generate schizophrenic behavior. *Neural Network*, URL: http://dx.doi.org/10.1016/j.neunet.2011.01.001.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947.

Montague, P. R., McClure, S. M., Baldwin, P. R., Phillips, P. E. M., Budygin, E. A., Stuber, G. D., et al. (2004). Dynamic gain control of dopamine delivery in freely moving animals. *Journal of Neuroscience*, *24*(7), 1754–1759.

Moutoussis, M., Bentall, R. P., El-Deredy, W., & Dayan, P. (2011). Bayesian modeling of jumping-to-conclusions bias in delusional patients. *Cognitive Neuropsychiatry* (in press).

Moutoussis, M., Bentall, R. P., Williams, J., & Dayan, P. (2008). A temporal difference account of avoidance learning. *Network*, *19*(2), 137–160.

Moutoussis, M., Orrell, M. W., & Morris, R. (2004). Modeling discoordination of cortical neuroactivity: relevance for the executive control of attention in Alzheimer's disease. *Journal of Integrative Neuroscience*, *3*(1), 85–104.

Murray, V., McKee, I., Miller, P. M., Young, D., Muir, W. J., Pelosi, A. J., et al. (2005). Dimensions and classes of psychosis in a population cohort: a four-class, four-dimension model of schizophrenia and affective psychoses. *Psychological Medicine*, *35*(4), 499–510.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl)*, *191*(3), 507–520.

Notturno, M. A. (1984). The Popper/Kuhn debate: truth and two faces of relativism. *Psychological Medicine*, *14*(2), 273–289.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–337.

Panksepp, J. (1998). *Affective neuroscience: the foundations of human and animal emotions*. New York: Oxford University Press.

Author's personal copy

Panlilio, L. V., Thorndike, E. B., & Schindler, C. W. (2007). Blocking of conditioning to a cocaine-paired stimulus: testing the hypothesis that cocaine perpetually produces a signal of larger-than-expected reward. *Pharmacology Biochemical Behaviour*, *86*(4), 774–777.

Ray, D., King-Casas, B., Montague, P. R., & Dayan, P. (2009). Bayesian model of behaviour in economic games. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems*: *Vol. 21* (pp. 1345–1352). MIT Press.

Redish, A. D. (2004). Addiction as a computational process gone awry. *Science*, *306*(5703), 1944–1977.

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: current research and theory* (pp. 64–99).

Rudy, Y., Ackerman, M. J., Bers, D. M., Clancy, C. E., Houser, S. R., London, B., et al. (2008). Systems approach to understanding electromechanical activity in the human heart: a national heart, lung, and blood institute workshop summary. *Circulation*, *118*(11), 1202–1211.

Sanislow, C. A., & Carson, R. C. (2001). Schizophrenia: a critical examination. In H. E. Adams (Ed.), *Comprehensive handbook of psychopathology* (pp. 403–441). Berlin: Springer.

Schmajuk, N., & Zanutto, B. (1997). Escape, avoidance and imitation: a neural network approach. *Adaptive Behavior*, *6*, 63–129.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience*, *13*(3), 900.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Smith, A., Becker, S., & Kapur, S. (2005). A computational model for the functional role of ventral-striatal d2 receptor in the expression of previously acquired behaviors. *Neural Computation*, *17*, 361–395.

Smith, A., Li, M., Becker, S., & Kapur, S. (2007). Linking animal models of psychosis to computational models of dopamine function. *Neuropsychopharmacology*, *32*(1), 54–66.

Soubrié, P. (1986). Reconciling the role of central serotonin neurons in human and animal behaviour. *Behavioral and Brain Sciences*, *9*, 319–364.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society B: Biological Sciences*, *216*(1253), 427–459.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press.

Tenenbaum, J., Griffiths, T., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318.

Tripp, G., & Wickens, J. R. (2008). Research review: dopamine transfer deficit: a neurobiological theory of altered reinforcement mechanisms in ADHD. *Journal of Child Psychology and Psychiatry*, *49*(7), 691–704. URL: http://dx.doi.org/10.1111/j.1469-7610.2007.01851.x.

Volkow, N. D., & Li, T. K. (2004). Drug addiction: the neurobiology of behaviour gone awry. *Nature Reviews Neuroscience*, *5*(12), 963–970.

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*(6842), 43–48.

Waltz, J. A., & Gold, J. M. (2007). Probabilistic reversal learning impairments in schizophrenia: further evidence of orbitofrontal dysfunction. *Schizophrenia Research*, *93*(1–3), 296–303.

Williams, J. (2006). Multiple timescales of evolution. *Behavioral and Brain Sciences*, *29*(4), 426–427.

Williams, J. (2008). Working toward a neurobiological account of adhd: commentary on gail tripp and jeff wickens, dopamine transfer deficit. *Journal of Child Psychology and Psychiatry*, *49*(7), 705–711. discussion 711. URL: http://dx.doi.org/10.1111/j.1469-7610.2008.01921.x.

Williams, J. (2010). Discounting and ADHD: multiple minor traits and states. In G. J. Madden, & W. K. Bickel (Eds.), *Impulsivity: theory, science, and neuroscience of discounting* (pp. 323–357). Washington, DC: APA Books.

Williams, J., & Dayan, P. (2005). Dopamine, learning, and impulsivity: a biological account of attention-deficit/hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology*, *15*(2), 160–79. discussion 157–159.

Williams, J., Sagvolden, G., Taylor, E., & Sagvolden, T. (2009). Dynamic behavioural changes in the spontaneously hyperactive rat: 3. Control by reinforcer rate changes and predictability. *Behavioural Brain Research*, *198*(2), 291–297. URL: http://dx.doi.org/10.1016/j.bbr.2008.08.046.

Williams, J., & Taylor, E. (2004). Dopamine appetite and cognitive impairment in attention deficit/hyperactivity disorder. *Neural Plasticity*, *11*(1–2), 115–132. URL: http://dx.doi.org/10.1155/NP.2004.115.

Willner, P. (1997). Validity, reliability and utility of the chronic mild stress model of depression: a 10-year review and evaluation. *Psychopharmacology*, *134*, 319–29.

Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. B., Hadden, S., & Burns, A. (1998). Health of the nation outcome scales (HoNOS). research and development. *British Journal of Psychiatry*, *172*, 11–18.

Wise, C. D., Berger, B. D., & Stein, L. (1972). Benzodiazepines: anxiety-reducing activity by reduction of serotonin turnover in the brain. *Science*, *177*(4044), 180–183.

World Health Organization, (1990). *International classification of diseases*. World Health Organization Press.

Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game theory of mind. *PLoS Computational Biology*, *4*(12), e1000254.

Yoshida, W., Dziobek, I., Kliemann, D., Heekeren, H. R., Friston, K. J., & Dolan, R. J. (2010). Cooperation and heterogeneity of the autistic mind. *Journal of Neuroscience*, *30*(26), 8815–8818.

Zaghloul, K. A., Blanco, J. A., Weidemann, C. T., McGill, K., Jaggi, J. L., Baltuch, G. H., et al. (2009). Human substantia nigra neurons encode unexpected financial rewards. *Science*, *323*(5920), 1496–1499. URL: http://dx.doi.org/10.1126/science.1167342.