Neuroscience and Biobehavioral Reviews 36 (2012) 2371–2379

Contents lists available at SciVerse ScienceDirect

Neuroscience and Biobehavioral Reviews

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

Commentary

# Empiricists are from Venus, modelers are from Mars: Reconciling experimental and computational approaches in cognitive neuroscience

Rosemary A. Cowell[a,*], Timothy J. Bussey[b,c], Lisa M. Saksida[b,c]

[a] Department of Psychology, University of California San Diego, 9500 Gilman Drive #0109, La Jolla, CA 92093-0109, USA
[b] Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK
[c] The MRC and Wellcome Trust Behavioural and Clinical Neuroscience Institute, Cambridge CB2 3EB, UK

ARTICLE INFO

*Article history:*
Received 18 October 2011
Received in revised form 16 August 2012
Accepted 23 August 2012

*Keywords:*
Computational model
Connectionism
Behavioral neuroscience
Cognitive neuroscience
Levels of organization
Problem space
Biological plausibility
Parsimony
Model parameters

ABSTRACT

We describe how computational models can be useful to cognitive and behavioral neuroscience, and discuss some guidelines for deciding whether a model is useful. We emphasize that because instantiating a cognitive theory as a computational model requires specification of an explicit mechanism for the function in question, it often produces clear and novel behavioral predictions to guide empirical research. However, computational modeling in cognitive and behavioral neuroscience remains somewhat rare, perhaps because of misconceptions concerning the use of computational models (in particular, connectionist models) in these fields. We highlight some common misconceptions, each of which relates to an aspect of computational models: the *problem space* of the model, the *level of biological organization* at which the model is formulated, and the importance (or not) of *biological plausibility*, *parsimony*, and *model parameters*. Careful consideration of these aspects of a model by empiricists, along with careful delineation of them by modelers, may facilitate communication between the two disciplines and promote the use of computational models for guiding cognitive and behavioral experiments.



## 1. Introduction

The statistician George E.P. Box famously declared that "all models are wrong, but some models are useful" (Box and Draper, 1987). In the cognitive and behavioral neurosciences – both relatively young disciplines in which few theories are well-established – this axiom is truer than ever. But it is precisely where well-established theory is lacking (where the models are most "wrong") that building models is most useful to scientific progress. In this review, we champion the use of concrete computational models in the search to understand the links between brain and behavior. We do so first by describing how such models can be useful when they are at their best, and second, by laying out a framework for deciding whether a model is useful in any given case.

In recent years, computational modeling has emerged as an extremely powerful tool in neuroscience. At a fine-grained biological scale, it can help to elucidate many diverse aspects of neural processing; a few recent examples include models that examine the characteristics of photon-sensitive ion channels (Foutz et al., 2012), the nature of temporal summation at transient receptor potential channels (Petersson et al., 2011), and the information processing advantages conferred by the different spiking and bursting modes of pyramidal cells in CA1 (Pissadaki et al., 2010). Low-level, realistic models of the information processing effected by neurons can provide critical insights into the consequences of known neurobiological details for the emergent properties of neural networks. However, computational modeling in neuroscience can be extremely valuable not just for low-level neuroscience, but also at higher levels of analysis where it can, for example, help to elucidate the links between brain and behavior. Some models that speak to behavioral phenomena are formulated at an intermediate level of analysis, employing neurobiological details such as realistic simulations of neural firing (Knight, 1972) and empirically observed mechanisms of neural plasticity such as long-term potentiation (LTP) (e.g., Sohal and Hasselmo, 2000; Bogacz et al., 2001) and adult neurogenesis (e.g., Aimone et al., 2009; Becker et al., 2009). Others are couched at a biological scale that is coarser still, a level which we will refer to as the 'anatomical systems' level. We use this term to describe models that provide an explanation of the function of a brain structure (or set of structures) that is defined by anatomical boundaries, such as perirhinal cortex (e.g., Cowell et al., 2006), or anterior cingulate cortex (e.g., Braver et al., 2001). In general, such models make fewer assumptions regarding low-level synaptic mechanisms or the details of neural processing, tending instead to focus on more abstract properties such as the type of representations contained in the brain region and the

* Corresponding author.
*E-mail address:* rcowell@ucsd.edu (R.A. Cowell).

0149-7634/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neubiorev.2012.08.008

mechanisms by which the region learns or processes the representations, associates them with certain responses, or makes decisions about them.

Across these different levels of theorizing, computational modeling is popular and prevalent in lower-level neuroscience, whereas more abstract (but still brain-based) computational models of cognition are less frequently used in partnership with empirical research, at the level of cognitive neuroscience. Yet, as we will suggest in this article, models formulated at the anatomical systems level can be as useful for theorizing about behavior as the lower-level models are for theorizing about neural information processing. Computational models of cognition and behavior facilitate the development of concrete hypotheses with well-specified assumptions. Model simulations then aid empirical investigations of behavior by distinguishing between viable and non-viable hypotheses ahead of time, narrowing the field of potential investigation. Moreover, instantiating a cognitive or behavioral theory as a computational model requires specification of an explicit mechanism for the function in question, which often produces clear, novel, and even counter-intuitive predictions for behavior. When simulations produce predictions that are novel or counter-intuitive, the model may be more easily distinguished from alternative, competing theories.

Despite these advantages, the symbiotic coupling of computational modeling and empirical behavioral neuroscience is exploited somewhat rarely. One possible reason for this is that empiricists can sometimes harbor certain misconceptions regarding computational models. These misconceptions can sometimes be caused by inadequate or incomplete communication of the modeling goals by modelers. But we believe that miscommunication between researchers from these two quite different academic disciplines need not be inevitable. We highlight some common misconceptions and note that each can be related to a particular aspect of a computational model, such as the identification of the problem space of the model, the level at which the model is couched, and the importance (or not) of biological plausibility, parsimony, and model parameters. Accordingly, careful and explicit consideration of these aspects of a model by empiricists, along with careful and explicit delineation of these aspects of a model by modelers, may facilitate productive communication between the two disciplines, promoting the use of computational models to guide experiments.

Broadly, this article will justify and advocate the use of *computational* modeling in cognitive and behavioral neuroscience. Of course, there are many successful and influential theories in this domain that are formulated in terms other than computational (e.g., verbal, diagrammatic); we do not propose that there is no place for them. Rather, we argue that computational theorizing is a useful method that is complementary to other forms of theorizing and, in particular, to empirical investigation (see also Cowell et al., 2011). We restrict our discussion largely to the set of computational models that can be described as connectionist models concerned with cognitive processes occurring in the brain. This restriction excludes two classes of neural network model: first, models within the domain of artificial intelligence that aim – legitimately – to build a useful machine without concern for whether the processes contained therein are related to cognitive mechanisms in biological organisms, and second, 'pure' psychological models that are unconcerned with how the algorithms they propose as a model of cognition might be implemented in biological tissue, for example, cognitive-level mathematical models and certain connectionist models (e.g., the model of word recognition and naming proposed by Seidenberg and Mcclelland (1989)). Our intended audience includes two groups of researchers: first, empiricists in the field of cognitive or behavioral neuroscience, and second, scientists with a specialized theoretical or mathematical background who would like to adapt their expertise to address empirical problems in an area of cognitive or behavioral neuroscience.

## 2. The field of cognitive neuroscience

Thanks to recent technological developments, cognitive neuroscience has at its disposal a wealth of powerful experimental tools for mapping function onto brain. The various techniques available for measuring brain function operate at a range of spatial resolutions (from patch clamp recordings to magnetoencephalography (MEG)), across a range of timescales (from single and multi-unit electrophysiology, through in vivo infusion of pharmacological agents, to naturally occurring lesions in humans), and offer a variety of different advantages, such as non-invasiveness combined with reasonable spatial resolution (e.g., functional magnetic resonance imaging (fMRI)), temporal precision combined with moderate cost (e.g., electroencephalography (EEG)), or sensitivity to specific variables of interest (e.g., in vivo monitoring of neurotransmitter levels through microdialysis). Moreover, in combination across a range of species, these techniques offer an unprecedented level of temporal and spatial resolution for the study of the circuitry and dynamics of the brain. With such a variety of techniques available, there may be a temptation to use them to generate as much data as possible, as quickly as possible. However, conducting empirical investigations of cognition from the stand-point of a well-specified theory can speed our acquisition of understanding immeasurably. Well-specified theories can give concrete predictions for experiments, are falsifiable and are explicit enough to be tested and refined, lending themselves to a systematic process of development into an ever more accurate model (Popper, 1999). The process of testing and refinement narrows down the number of experiments required in the investigation, focuses the research direction and encourages thorough, mechanistic understanding of the underlying processes. This theory-driven approach contrasts with one in which data is generated in a more piecemeal fashion by testing stand-alone hypotheses that do not form part of a comprehensive theory. Hypotheses that may be described as 'stand-alone' have typically been derived intuitively according to prior empirical results, rather than according to a mechanistic account of cognition that offers an explanation for the observed effect (and not simply a description of it). Testing such stand-alone hypotheses might guide us toward a more detailed characterization of a cognitive process or a deficit in a patient population; in this way, a data-driven approach can advance our intuitive understanding and may suggest important improvements in clinical treatment or diagnoses. However, progress in truly understanding a cognitive process or clinical deficit is likely to be incremental unless the hypothesis is part of a fully developed theory that links brain to behavior in a mechanistic way.

That is not to say that a bottom-up approach to investigating cognition should never be taken. Indeed, there are persuasive arguments in favor of an interactive or co-evolutionary strategy, which combines bottom-up (data-driven) and top-down (theory-driven) research strategies (e.g., Churchland and Sejnowski, 1992; McClelland et al., 2010). Top-down, or 'function-first', strategies begin with abstract principles concerning cognitive functions and only later attempt to map these principles onto neural processes (Griffiths et al., 2010). Bottom-up, or 'mechanism-first', strategies begin with knowledge of the mechanisms implemented in neurons, and use them to generate theories of psychological phenomena, for example, figure ground segregation or selective visual attention (Koch, 1993). A co-evolutionary research strategy can combine advantages from each, such as the tendency of a data-driven approach to generate new ideas about cognitive mechanisms

through the discovery of intricate or counter-intuitive biological detail, and the capacity of a top-down approach to guide empirical research in a focused manner. An interactive research strategy is often adopted in connectionism, by exploiting principles of information processing consistent with biological properties of the brain in the search for the abstract, algorithmic processes underlying cognition and behavior (Mcclelland and Plaut, 1993; McClelland et al., 2010).

## 3. The approach of computational modelers

Computational modeling as a tool for theorizing in cognitive neuroscience has many advantages, which have been eloquently described by many authors (Sejnowski et al., 1988; Mcclelland and Plaut, 1993; O'Reilly, 1998; McClelland et al., 2010). We summarize those advantages briefly, here. Although models from a variety of traditions possess the desirable properties we discuss here, we focus mainly upon connectionist models in this review because they are of particular utility to cognitive and behavioral neuroscience.

First, instantiating a theory in a computational model forces the theorist to specify the details of that theory; in writing a computer program, the assumptions, representations and operations of the theory are given as concrete statements. Once written, the theory is rigorously laid down and fully specified.

Second, computational modeling can reveal underlying complexities and unexpected predictions of a theory that may not be realized without simulations. At the outset of building a computational model, the theorist has likely considered the assumptions underlying the theory and should therefore have a good idea of the predictions that will emerge. However, there may be additional, unappreciated assumptions in the verbally or diagrammatically outlined theory that are revealed only when the code is written. Moreover, the consequences of certain assumptions (whether those assumptions were apparent from the outset or not) might be difficult to determine a priori, particularly when those consequences are the results of complex dynamics involving distributed representations. Allowing a simulation to do the number-crunching that follows from the theory can reveal consequences of the theory's assumptions, in terms of predictions for behavior, that the theorist him- or herself might not have anticipated. That is, the complexity of a model can block the model builder from appreciating the consequences of its underlying assumptions, but running simulations can reveal them, helping us to understand our own theories more fully. An example of this is given by Huber et al. (2002), who tested novel and unexpected predictions of a computational model of temporal perception. The model, first presented in Huber et al. (2001), accounted for short-term priming effects, in which the brief presentation of a stimulus before its subsequent appearance as a to-be-detected target either caused performance to improve (if the prime was presented only briefly) or to decline (if the prime lasted longer). The model accounted for this phenomenon by assuming that an observer can become confused about the source of incoming information when events are closely spaced in time, but can also implement 'discounting' of evidence for sources of perceptual information that are known to have recently been active, to counteract the effects of source confusion. The model is a probabilistic Bayesian account of temporal perceptual processing, but it operates upon stimulus representations that are "distributed" in a connectionist sense, capturing important stimulus properties such as feature-based similarity. In running probabilistic simulations using thousands of examples of feature-based stimulus representations, the authors revealed important limitations to the interaction between the two mechanisms that were assumed to underlie temporal perception: 'source confusion' and 'discounting'. In sum, although the theoretical framework seemed at the outset to imply that source confusion could always be offset by discounting when the prime stimulus was presented for a long duration, this was discovered not to be the case. Simulations predicted that if the *target* stimulus was presented very briefly, or if the target and prime stimuli were perceptually weakly related, discounting was unable to counteract the effects of source confusion. These predictions were critical in setting apart the model of Huber et al. (2001) from a competing theory of short-term priming authored by Ratcliff and McKoon (2001), and were confirmed empirically in Huber et al. (2002). Interestingly, because this model employed distributed, connectionist-like representations it lent itself naturally to a neural network implementation, which tied the source-confusion and discounting hypotheses to neural mechanisms, linking its account of behavior to the brain (Huber and O'Reilly, 2003).

A third advantage of computational models, particularly those in the connectionist tradition, is that they are necessarily mechanistic. A connectionist model requires representations of both inputs and outputs and a set of operations to transform those inputs into outputs, all of which are formulated according to the hypothesized mechanism underlying cognition. Some verbal theories might be subject to the criticism that their assumptions and operations are slippery because they are not mathematically specified: the verbal description of a cognitive process can mean different things in different situations. This can lead to excessive flexibility of a model in accounting for data and may render it unfalsifiable. Computational models, though not foolproof in this regard, are less likely to be excessively flexible because they are constrained by a concrete mechanism that is instantiated in computational code. Other verbal theories might give the impression of mechanistic explanation when in fact they are just descriptive. Computational models can help to flesh out such verbal theories, providing a mechanism for the processes described. The model of facial expression recognition presented by Dailey et al. (2002) constitutes a good example of this. There existed a verbally stated theory of facial expression recognition (Etcoff and Magee, 1992) that proposed that expressions are subject to "categorical perception" (CP), that is, the perceptual mechanisms underlying expression recognition are tuned to sharply defined emotion categories such that any given stimulus is mandatorily assigned to one or other category. The EMPATH model of Dailey et al. showed precisely how CP might emerge from cortical representations, by implementing the task with a pattern classifier that incorporated biologically plausible representations of visual stimuli.

A fourth advantage specific to connectionist computational models is that they commonly employ distributed representations, in which stimuli that share behaviorally relevant properties typically elicit activation across overlapping subsets of units in the network. This can give rise to interesting and useful properties of representations, such as generalization of learned responses to novel items, when those items share properties with stimuli that have been experienced previously (Hinton et al., 1986). In addition, the use of distributed representations in computational modeling is likely to aid convergence upon a theory in which the information processing operates in a brain-like manner, since there is evidence that distributed representations are used by biological neural networks in the hippocampus and other cortical areas (Zhang et al., 1998; Kilgard and Merzenich, 1999).

## 4. Combining the two approaches

Considering the advantages of conducting empirical research that is theory-driven, and the advantages of computational modeling as a theoretical tool, the combination of empirical behavioral

neuroscience and modeling should provide an extremely fruitful research approach. An example of this profitable symbiosis is provided by an investigation of anterior cingulate cortex (ACC) in speeded response tasks by Jones et al. (2002). A 'conflict-monitoring' theory of ACC function was instantiated in a connectionist network that successfully simulated behavioral performance on three different tasks, as well as fMRI activation of the ACC during performance of each task (Braver et al., 2001). The model hypothesized a functional role of 'conflict detection' in speeded response tasks, and drew a convincing link between that cognitive process and the activation of the ACC. In addition, prior to this modeling study there existed an apparent discrepancy between the brain imaging and behavioral data that had been difficult to reconcile with previous instantiations of the ACC conflict-monitoring hypothesis: although ACC activity was equivalent across three different speeded response task paradigms (2-alternative forced choice (2AFC), go/no-go and oddball), modulations in accuracy by target-frequency were seen in only the two-response speeded response task paradigm (2AFC), and not in either of the one-response speeded response task paradigms (go/no-go and oddball). The model reconciled these findings by demonstrating a plausible mechanism whereby conflict-monitoring in ACC plays a critical role in task performance for all three paradigms, but differences in overt behavior between two- versus one-response speeded response task paradigms are generated downstream of conflict detection. Thus, the use of computational modeling helped to understand counter-intuitive effects in the empirical data, and in turn, the empirical data were used to refine the theory.

If the fields of cognitive neuroscience and connectionism can be combined so effectively, why is this combination not more often seen? Largely, we would argue, because of a problem with communication.

In order to iron out misunderstandings between the model builder and the critic, we must decide on the expectations that each party can reasonably hold. On the one hand, what can cognitive and behavioral neuroscientists evaluating a model reasonably demand, and what should they not expect? On the other hand, in which ways is it critical for modelers to define the scope and aims of their model, and how best should they communicate these properties? Our attempt to facilitate communication between these two different fields is in the spirit of Churchland and Sejnowski (1992). We will address five areas of potential miscommunication: the level of biological organization at which the model attempts to address data, the problem space of the model, the biological plausibility of the model, the issue of parsimony, and the importance of parameters. For each area, we will describe how misconceptions can arise. In each case, we will outline how we believe computational modelers must clearly define their position in order to avoid such misconceptions, and highlight some considerations for the appropriate evaluation of computational models by both empiricists and modelers alike. In doing so, we hope to reveal a set of guidelines that may be used for deciding whether a given model is useful, or too wrong to be useful.

## 5. Five critical properties of computational models

### 5.1. Levels of biological organization

In neuroscience, the level of biological organization of a theory refers to the scale of the biological components that feature in the explanation that the theory offers. For complex organisms, there exists a hierarchy of levels of biological organization that emerges from the structure of the organism itself: an individual animal possesses many organs, including a brain; brains are composed of lobes and anatomically distinct systems (e.g., the parietal lobe, or the hippocampal system); each anatomical system contains networks of neurons; neurons contain axons, cell bodies, ion channels and so forth; these components in turn carry out their function via molecular interactions at a still smaller scale. Churchland and Sejnowski (1988) discuss the levels of organized structure at which research can be conducted in neuroscience, ranging from molecular interactions within and between cells to mechanisms at the level of the central nervous system (CNS) as a whole (see Fig. 1).

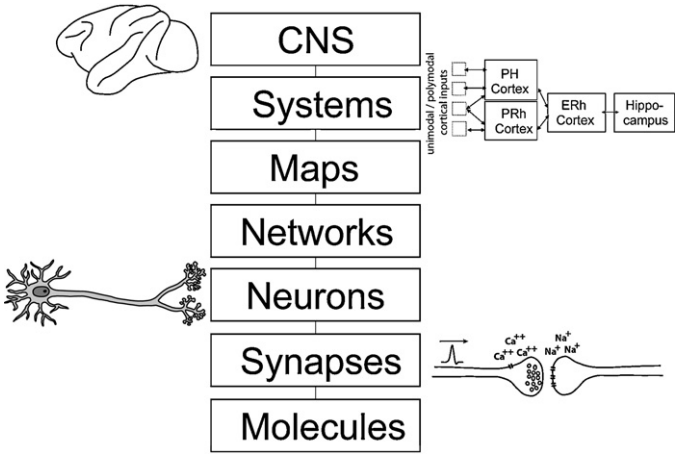


**Fig. 1.** Levels of biological organization. Schematic illustration of the levels of biological organization at which a problem may be studied. After Churchland and Sejnowski (1988).
Adapted from Cowell et al. (2011).

The mechanisms of cognition may be investigated at any of the levels of biological organization shown in Fig. 1. Often, explanations of a single cognitive phenomenon can exist simultaneously at several levels of biological organization. For example, memory might be explained both at the level of individual synapses – for example in terms of LTP – and at the level of brain systems – in terms of the multiple interacting anatomical systems (e.g., visual and auditory cortices, hippocampus and prefrontal cortex) that contribute to the encoding of a rich episodic memory trace. Moreover – unlike Marr's *levels of analysis* (Marr, 1982), which he argued to be formally independent – the mechanisms that operate at each level of biological organization are far from independent; for example, the manner in which neurons within a network interact can depend critically on the types of synapses those neurons possess. Given that the different levels in the hierarchy can interact, and that cognitive phenomena can be understood at multiple levels, neuroscientific theories can sometimes be usefully couched across a small range of biological levels rather than strictly adhering to a single level. Either way, it is vital that both modelers and evaluators of models consider the hierarchy of organizational levels carefully: knowing *where* in the hierarchy of levels a model of cognition is situated is fundamental to the proper understanding and evaluation of the explanation offered by that model.

Misunderstanding of the organizational level at which a model is pitched can lead to inappropriate demands on a model. Those appraising computational models need to be especially mindful in deciding when it is appropriate to forgive simplifying assumptions of the independence of levels (e.g., modeling a cognitive phenomenon at the systems level without including detail from the synaptic level), and when it is appropriate to insist on interaction between levels. A hypothetical example is given by the connectionist model of Ashby et al. (1998), COVIS, which offers a brain-based account of category learning. This model is formulated at the anatomical-systems level. It possesses several components that perform distinct cognitive goals such as 'computing a high-level visual representation', 'associating a stimulus with a category response' or 'rule selection', and each component corresponds to

an anatomically defined brain region, such as extrastriate (inferior temporal, IT) cortex, striatum or anterior cingulate cortex. The model makes predictions for category learning in both healthy participants and patient populations. The architecture and the chain of cognitive events assumed by the model are strongly inspired and supported by neuropsychological and neuroanatomical data, but the model does not attempt to account for low-level neurobiological details such as neural firing patterns. For example, no attempt is made to explain how the visual representations are constructed by neurons in IT cortex, because the primary aim of the model is to investigate how stimulus representations in this region (once constructed) participate in category learning through interaction other brain regions. The fact that stimulus representations exist in IT cortex is well known, and their construction is the subject of a great many other models addressing the equally formidable problem of object recognition (e.g., Fukushima, 1980; Wallis and Rolls, 1997; Dailey and Cottrell, 1999; Riesenhuber and Poggio, 1999; Serre et al., 2007), hence COVIS justifiably assumes this stage of processing to be complete. It would therefore be inappropriate to insist that the COVIS model include low-level biological details such as the repetition-sensitive responses of IT neurons (e.g., Miller et al., 1991) in its mechanism. Repetition-sensitive responding is a neuron-level mechanism, and COVIS is couched at the anatomical-systems level. In this hypothetical example, such an inappropriate demand might stem from one of two sources. Either the model evaluator has misunderstood the organizational level at which the model is couched, or they have understood the level of the model but are insisting that the model should additionally account for data from a different level and simulate the interaction between those levels. In the second case, the error would lie in premature insistence upon a multi-level explanation: when a problem remains poorly understood at the anatomical systems level alone, it is acceptable to model only that level of biological organization in an attempt to account for unexplained data. Indeed, building a simple, clear theory that accounts for the anatomical-level data is an important and necessary first step in scenarios such as this. Another reason for excluding lower-level details (such as repetition sensitive responses of neurons) is that they are not a necessary part of the mechanism by which the behavior (category learning) is explained. To include unnecessary computational detail can obscure the key assumptions driving the critical behavioral effects.

The risk of such misunderstandings arising is greatly reduced if the modeler takes care to make clear exactly at which level, or levels, of biological organization the explanation of the model is aimed. If the model spans more than one level, the author must specify whether it uses data from lower levels of organization according to strict constraints, or simply as an inspirational guide to the kind of processing that might operate at a higher level. Choice of the level of biological organization of a model is also relevant to the idea of problem space, which we discuss next.

### 5.2. Problem space

The problem space of a model defines the intended scope of the model's explanatory power. In neuroscience, it can be thought of as a two-dimensional space, as illustrated schematically in Fig. 2. The two dimensions are the number of phenomena (in the present discussion, cognitive phenomena, e.g., visual discrimination, recognition memory, spatial attention) that a model attempts to explain and the number of levels of organization at which it attempts to explain them.

Problem space can be constrained enormously by choosing a particular scale, or level of biological organization, at which to build the model. Constraining a model to one or a few biological levels is advantageous because details from other levels may be irrelevant and would reduce the clarity of the model's account.

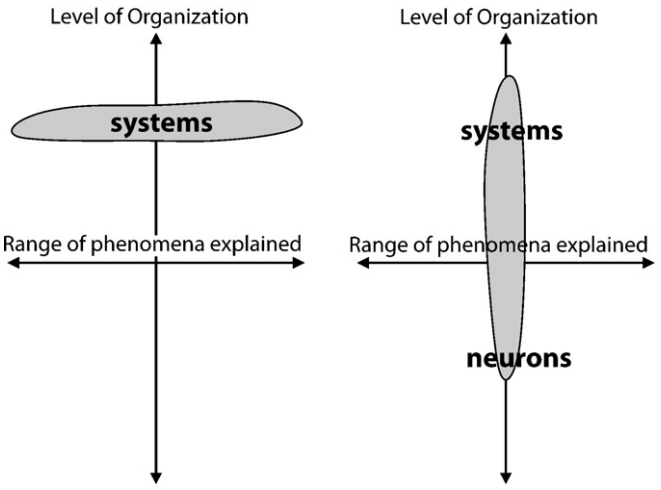


**Fig. 2.** Problem space. Two illustrative examples of the problem space that may be adopted by a computational model. Left: a model that attempts to deal with many phenomena (e.g., categorization, priming and recognition memory) at only one level of biological organization – systems. Right: a model that addresses a very specific cognitive phenomenon but attempts to explain its mechanism at many levels of biological organization.
Adapted from Cowell et al. (2011).

Building a complex model that includes all known biological detail can lead to the simulation being as poorly understood as the nervous system itself (Sejnowski et al., 1988). This principle applies to many branches of science: a useful meteorological model that makes weather forecasts or predicts the formation of air pollutants does not need to include the energetic states of individual water molecules in the atmosphere. However, the choice of one biological level need not preclude the model from being *sensitive* to the levels below and above; drawing inspiration or constraints from details of nearby levels can greatly facilitate development of the hypothesized mechanisms at the chosen level. A good example of this is offered by ACT-R, a framework that was originally proposed as a purely cognitive architecture (Anderson, 1993); that is, it initially explained behavior at the abstract, psychological level without reference to biology. However, recent development of ACT-R has exploited fMRI, examining the activation of different brain regions (defined at the anatomical systems level, e.g., motor cortex, fusiform gyrus and anterior cingulate cortex) to assess the validity of ACT-R's assumptions about the partitioning of cognitive processes into separate modules (Anderson, 2007; Anderson et al., 2008). In this way, Anderson and colleagues have directed the development of a cognitive-level theory of problem solving (and more generally, a theory of cognitive architecture) by using brain imaging to reveal the properties of a lower level – the anatomical systems level – in order to constrain the theory.

In contrast to the need for constraint over levels of organization, it is generally advantageous for a model to account for as many related phenomena (at the same level of organization) as possible. However, in order to define a computationally tractable problem, constraining the problem space to a limited number of phenomena may be necessary. The pursuit of too large a problem space can produce a model that is not at all parsimonious. Additionally, the problems in cognitive neuroscience addressed by connectionist models are often poorly understood; attempting to model several phenomena might reduce the clarity of the key demonstration made by the model, or might simply prove prohibitively ambitious. To provide a concrete example, consider the construction of a model of a very specific aspect of cognition such as visual recognition memory. Such a model needs to explain how an animal can make a judgment of familiarity about an object that

is presented to it. This explanation can be achieved by simulating the development of a stimulus representation with visual experience so that it looks familiar the second time the network sees it. The model need not also explain everything else that an animal does in a recognition memory task, such as deciding that the peanut reward is worth performing for, planning an arm movement, executing the motor command to reach out and indicate that the object is familiar, and so on. To model all of these behavioral feats would require a theory of vast scope, incorporating memory, motivation and executive control. It is thus justifiable to focus on only one aspect of behavior, simplifying or assuming all other aspects of the behavioral phenomenon not within the pre-defined problem space.

The delineation of a clearly defined problem space at the outset of model building is an extremely important step: without it, problem space, like levels of biological organization, can be an area where misconceptions arise. In turn, to avoid such misconceptions, the critic of a connectionist model must consider whether his or her criticism is appropriate for the explanation being offered by the model. For example, it would not be sensible to criticize a model of the role of the hippocampus in spatial memory because it does not explain the hyperactivity seen in animals with hippocampal lesions, if the author of the model has made it clear that the behavior that the model is intended to explain is spatial memory. (Provided, of course, that the model does not make *incorrect* predictions regarding activity levels. It is acceptable for the model to make no predictions at all about activity levels, but in the unlikely event that the mechanism accounting for spatial memory entailed an inescapable prediction for activity levels that contradicted the observed increase in activity caused by hippocampal lesions, the model should of course be deemed incorrect.) Indeed, hippocampal function offers a second example of how a model's problem space must sometimes be limited: most brain-based models of episodic memory are partly or wholly localized to the hippocampus (e.g., McClelland et al., 1995; Norman and O'Reilly, 2003; Meeter et al., 2005; Greve et al., 2010) and yet many such models do not address the well-documented role of hippocampus in spatial memory or navigation. Because there are such a wide variety of datasets and empirical phenomena pertaining to episodic memory function alone, addressing both episodic and spatial memory with a single computational account may often be too ambitious.

But not all examples are so clear cut. Another is offered by a model of visual discrimination learning proposed by Bussey and Saksida (2002). The authors clearly stated the problem space to be that of visual discrimination – learning or remembering how to tell two visual stimuli apart from one another. But a critic might argue that the same model ought to account for delay-dependent deficits in object recognition, since this is the canonical deficit observed following lesions in perirhinal cortex. The critic has a good point; a more comprehensive theory should aim to account for these data, and a justification should be offered for why the model does not address them. The problem space of this model was not expanded to deal with the phenomenon of object recognition in the first instance for two reasons. First, the behavioral data on discrimination learning were complex and puzzling enough alone, so that a model of just these effects would constitute a significant advancement in understanding. Second, the ideas that were under development regarding the explanation of object recognition memory deficits seemed likely to hinge upon the explanation of simple visual discrimination. The additional simulation of the effect of a retention delay (as used in object recognition memory tasks) is therefore a mechanism that would be explained by building on top of the discrimination model (which was later done, see Cowell et al., 2006). The discrimination model therefore had to be well worked out and tested first. Thus, in this case, restricting the number of phenomena modeled seems justified, but this example illustrates the shades of gray that exist in making such restrictions. In particular, the size of the problem space enters into a trade-off with other factors such as parsimony, as is discussed below. Importantly, just as with levels of biological organization, it is up to the modeler carefully to delineate and justify the problem space, particularly if there is danger of it appearing narrow to the skeptic.

### 5.3. Biological plausibility

All brain-based models of cognition necessarily strive for some degree of biological plausibility. O'Reilly (1998) argues persuasively in favor of biological plausibility: "Biological realism lies at the foundation of the entire enterprise of computational modeling in cognitive neuroscience. This approach seeks to understand how the brain... gives rise to cognition, not how some abstraction of uncertain validity does so. Thus, wherever possible, computational models should be constrained and informed by biological properties of the cortex. Moreover, computational mechanisms that violate known biological properties should not be relied upon." This philosophy provides a useful guiding principle for researchers interested in brain-based cognition. However, within the realm of biologically plausible models, there are at least two distinct subclasses between which it is useful to make a distinction: realistic models and simplifying models.

Rolls and Deco (2002) advocate a realist approach, arguing that while connectionist approaches make an important start on understanding how complex computations such as language could be implemented in brain-like systems, if the model uses, for example, back-propagation or too few neurons, it can only provide a guide as to how cognition might be implemented in the brain. In its most extreme form, the realist tradition builds models that operate according to strict rules of biological plausibility and incorporate as many low-level details as possible (Sejnowski et al., 1988). We would argue that there are in fact several disadvantages to using realistic models in the domain of cognitive neuroscience. First, realistic models cannot feasibly be pitched at a level of any organization any higher than the neuron or network level, since incorporation of cellular detail in a model of systems-level processes would result in a model so complex as to make analysis of its mechanisms intractable. Second, historical precedent suggests that some proportion of experimentally established facts are likely to be revealed as wrong or inaccurate in detail as scientific progress is made: this tendency may have a greater negative effect on detailed models than on simpler ones. Finally, it is unclear where the adherence to known biology should stop: is it sufficient to employ biologically plausible synaptic updating rules, without implementing individual inhibitory and excitatory postsynaptic potentials, or realistic placement of the ion channels within the membranes? Returning to George Box, "all models are wrong" – no matter how hard one labors to include all known biological detail, a perfect simulation of the brain will never be achieved. Of course, achieving such a thing would amount to synthesizing a real brain, and if we could do this, our understanding of the brain would be complete: we would no longer need models. It is our incomplete understanding that necessitates models, which allow us to simplify reality in order to understand it. Models are thus by definition wrong (because they omit information), but in some sense their utility depends upon it. That is, by attempting to include all known biological detail, the modeler would be prevented from focusing upon the aspects of the neural mechanism that are assumed to account for the effect of interest. And it is by focusing on these key mechanistic aspects that we can properly test our hypotheses, by ensuring that the model predictions do indeed stem from the critical part of the hypothesized mechanism.

Realistic models contrast with a connectionist approach to cognitive neuroscience, in which models aim to simplify the problem under investigation. When seeking an explanation at the

systems-level or higher – as is often the case in cognitive and behavioral neuroscience – connectionist models, which capture the important principles underlying brain function without incorporating all known neurobiological detail, are more useful than realistic models. Most connectionist models implement only the key mechanism of interest in order to simulate behavior. By deliberately paring down the problem to a clear and comprehensible mechanism, eschewing biological details that are not a necessary part of the account, the hypothesized mechanisms are subject to the most stringent test of their ability to explain the target data. Thus, simplifying models can be used to test critical assumptions regarding which emergent properties of low-level mechanisms are important for producing behavior at the level of the whole organism. As cautioned by Churchland and Sejnowski (1992), critics of connectionism should not assume that a high degree of realism always equates to a high degree of scientific value.

An example of a common misunderstanding regarding the biological plausibility of computational models is provided by the back-propagation algorithm. The criticism frequently arises that a connectionist model can be disregarded as biologically implausible if it uses back-propagation, since it is thought unlikely that the back-propagation of error could be implemented locally in biological neurons (e.g., Crick, 1989). If the model is couched at the *systems level* this criticism is inappropriate and likely stems from confusion over the model's level of biological organization: back-propagation is a neuron- or network-level mechanism, and so theories at the systems level need not be concerned with the biological plausibility of back-propagation. It is possible that large ensembles of neurons acting in concert at the network level could indeed effect some kind of error-correcting learning process that would produce results at a systems level that are akin to the systems-level mechanism being proposed, so the model need not be dismissed wholesale on the ground that it could never be implemented in the brain. For example, Gluck et al. (2001) have discovered a circuit-level mechanism of learning in the cerebellum and brain stem that can account for the error-correcting property of the Rescorla–Wagner rule. They state that "[the mechanism] is an emergent property of the organization of the neural circuit itself rather than a specialized synaptic process". And, critically, a systems-level model need not concern itself with the details of that network-level process. Moreover, when attempting to understand the nature of a cognitive process at the systems level, the structure learned by a model employing back-propagation (or other biologically implausible mechanism) may be interesting and plausible, even if the learning algorithm itself is not; this holds particularly true in the light of the findings from Gluck et al. (2001).

However, if one tried to devise a model at the *network level* that relied on back-propagation, then biological plausibility of the algorithm becomes a serious constraint. Similarly, a systems-level model that relied on direct connections between structure X and structure Y would be unworkable if such direct connections are known not to exist, since the connectivity of brain structures is an aspect of biological plausibility with which a systems-level model must be concerned.

### 5.4. Parsimony

Following Occam's Razor, a scientific model should aim to provide the simplest version of events that can account for the evidence. For investigators of artificial intelligence, as well as purely cognitive theorists who have no interest in the hardware used to implement the algorithms that they seek to understand, parsimony should always be a priority. Parsimony is also a sensible guiding principle when engineering the most efficient machine. But the task of the cognitive neuroscientist is instead one of reverse engineering: we are interested in how the evolving human brain came upon solutions to the problems presented to our biological ancestors, via natural selection. The process of natural selection does not, of course, have an a priori purpose to build the most efficient machine, with the end goal of modern humans in its sights; humans and other animals are the product of an evolutionary process in which our cognitive ability has been augmented in a step-wise manner by building on pre-existing structures. This process does not necessarily yield the most elegant computational solutions, and so we must be open to possible absences of parsimony in the brain's design. Cognitive neuroscientists should therefore be prepared to disregard Occam's Razor in the rare cases where the neurobiological or behavioral evidence unambiguously counters the simplest possible explanation.

Nonetheless, a theorist must have a good, empirically justified reason to construct a model that is more complex than is necessary to explain the primary target data. It is inappropriate to insist upon the inclusion of biological facts not relevant to the phenomenon being modeled and unnecessary to the proposed mechanism for cognition. In fact, the pitfall of over-adherence to biological detail illustrates the intimate links between parsimony, problem space, levels of biological organization and biological plausibility. Inappropriate insistence upon adherence to biological detail can arise in some cases from failing to appreciate the defined problem space of the model in terms of the level of biological organization at which the model is formulated. In other cases it might stem from confusing the need to ensure that there is no evidence indicating that a systems-level model *could not* be implemented in biological tissue, with the need to demonstrate how a lower-level model *likely is* implemented in biological tissue. The interaction of these issues underlines the importance of clearly defining (for model builders) and fully appreciating (for model evaluators) the model's intended problem space. There exists a trade-off between the parsimony of a model and the size of its problem space, which is intimately linked to the level of biological organization of the model and the constraints imposed by biological plausibility at that level.

### 5.5. Parameters

One final area we will discuss, in which misconceptions concerning computational models can arise, concerns the parameters and output of the models. A perhaps counter-intuitive heuristic with regard to computational models in cognitive and behavioral neuroscience is that the absolute values of the simulation results produced by a model, and their variability, are often unimportant. In simulating a given task, a computer program might take 200 trials to learn a problem and a rat 54. This result is hardly surprising, since the model is likely addressing only one, or a very few, aspects of behavior (e.g., a computational model of reversal learning may not model attention, decision making, locomotion or motivation in the animal performing the task). The key property of the simulations is the qualitative trend or trends in the data that emerge, for example in comparing a lesioned version of the model with an intact version to simulate the effects of brain damage, or comparing the model's performance on different task conditions. If the trends in the simulation data are qualitatively the same as the trends in the behavioral data, the model is valuable because its mechanism may be the correct one for explaining the trends in the behavioral data.

In support of this argument, consider the use of different animal species to study the same behavioral phenomenon. Many behavioral tasks have been adapted for use across species and are deemed analogous if the trends that emerge are the same; we would not necessarily reject an analogy between tasks measuring rat and monkey reversal learning because the rat takes longer to learn it than the monkey does. The rat has entirely different physical and cognitive parameters, in terms of locomotion,

motivation, attention, decision making and perhaps even reversal learning itself. The discrepant learning rates may arise because of these different parameters, rather than because the task measures something qualitatively different in the two species. Indeed, in this real-world example, despite the differences in acquisition rates it has been determined that reversal learning in mice, rats, monkeys and humans relies on similar brain regions and neurotransmitter systems, for example the orbitofrontal cortex (e.g., Iversen and Mishkin, 1970; Butters et al., 1973; Birrell and Brown, 2000; Schoenbaum et al., 2002; McAlonan and Brown, 2003; Izquierdo and Murray, 2004; Izquierdo et al., 2004; Boulougouris et al., 2007; Tait and Brown, 2007; Bissonette et al., 2008; Rudebeck and Murray, 2008; Boulougouris and Robbins, 2009; Hampshire et al., 2012). Just as qualitatively similar findings across different species have been interpreted convincingly within the same framework, comparison between qualitative patterns of data from computational models and animal experiments can be appropriate, regardless of quantitative identity of the outputs.

Of course, sometimes the results of computational models do provide quantitative predictions, and here the numerical values emerging from the models may be important. For example, Nosofsky et al. (1992) compared three models of category learning on their accuracy in simulating empirical data. All three models produced an output value corresponding to the probability of correct classification, which varied from 0 to 1 in all simulations. Behavioral data from human subjects performing categorization tasks were also recorded as probability of correct classification. The authors were therefore able to compare the models' performance not only on their ability to reproduce qualitative trends in the data, but also on their quantitative fit to the empirical data. This analysis allowed a more sensitive model comparison than would have been possible by considering only qualitative trends, since the models produced qualitatively similar results in many of the tasks simulated. Thus, in evaluating the output of a computational model, we must first consider whether it is the quantitative or qualitative trends (or both) that are important.

Another criticism of connectionist models often heard is that modelers are simply "playing the parameter game". Connectionist models have many degrees of freedom, both in the architecture of the model and in other parameters that determine the learning processes within the model; critics argue that, given sufficient tweaking of the parameters, any behavioral phenomenon could be modeled, so the eventual discovery of a solution with one particular set of parameters is inevitable and therefore uninteresting (see also O'Reilly and Munakata, 2000). However, this tweaking process becomes less and less powerful (or the degrees of freedom of the model become fewer) as the problem space is enlarged. If a model is required simultaneously to account for a number of behavioral phenomena across more than one subject group (e.g., brain damaged and intact), and if the model makes novel predictions that are tested experimentally, it becomes less likely that any arbitrary mechanism would be successful, given the right set of parameters. This argument requires, however, that the same set of model parameters is used across all of the phenomena, task conditions and subject groups modeled. An example of a model whose mechanism is rendered convincing by the sheer number of phenomena accounted for is given by Bowman and Wyble (2007), who present a connectionist model of temporal attention and working memory that accounts for the well-documented "attentional blink" (AB). The AB is observed when a participant views a rapid sequence of visual stimuli presented at the same spatial location, and is required to detect certain target items within the stream. Typically, when two target items are presented in succession, the participant will fail to detect the second item if it appears within 200–500 ms after the first (Raymond et al., 1992). The AB provides a window into the mechanisms of temporal attention and has thus been extensively investigated. Bowman and Wyble accounted for seven documented empirical phenomena related to the AB with their model, and made three further, novel predictions, which were tested and confirmed with new experiments. The model does not produce a perfect quantitative fit to behavioral data for every simulated phenomenon. But the impressive breadth of empirical trends qualitatively reproduced mitigates any potential concern that the mechanism is an arbitrary one that was made to fit the data via excessive parameter tweaking. Rather, because so many empirically observed phenomena emerge from this mechanism it seems likely to have captured at least some key properties of temporal cognition. This consideration renders it useful in advancing our understanding of temporal attention.

## 6. Conclusions

We believe that much can be done to facilitate the use of computational models in cognitive and behavioral neuroscience. First, it is the responsibility of the modeler to clearly define, for every model constructed: the level of biological organization at which the model is couched; the problem space of the model; the degree of biological plausibility, or realism, for which the model strives; the importance or otherwise of parameters and absolute numerical values to the model's account; and, lastly, where sacrifices have been made in any of these areas for the sake of parsimony, or if there is a deliberate lack of parsimony in line with biological constraints. These definitions are critical to ensuring the model is comprehensible and valuable to empirical neuroscientists. It is then the responsibility of the model appraiser to take into consideration the five key issues discussed, in order to ensure constructive and reasonable criticisms of brain-based computational models of cognition. Under conditions of good communication and mutual understanding, modelers and empiricists should between them be able to decide which models are useful and which are too wrong to be useful. In this way, effective communication between empiricist and theorist can be achieved, so that the two can work in tandem toward a better understanding of the brain basis of cognition.

## Acknowledgements

We thank David Huber for helpful comments on an earlier draft of the manuscript. RAC was funded in part by NSF grant BCS-0843773.

## References

Aimone, J.B., Wiles, J., Gage, F.H., 2009. Computational influence of adult neurogenesis on memory encoding. Neuron 61, 187–202.
Anderson, J.R., 1993. Rules of the Mind. Erlbaum, Hillsdale, NJ.
Anderson, J.R., 2007. Using brain imaging to guide the development of a cognitive architecture. In: Gray, W.D. (Ed.), Integrated Models of Cognitive Systems. Oxford University Press, New York, NY, pp. 49–62.
Anderson, J.R., Carter, C.S., Fincham, J.M., Qin, Y., Ravizza, S.M., Rosenberg-Lee, M., 2008. Using FMRI to test models of complex cognition. Cognitive Science 32, 1323–1348.
Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., Waldron, E.M., 1998. A neuropsychological theory of multiple systems in category learning. Psychological Review 105, 442–481.
Becker, S., MacQueen, G., Wojtowicz, J.M., 2009. Computational modeling and empirical studies of hippocampal neurogenesis-dependent memory: effects of interference, stress and depression. Brain Research 1299, 45–54.
Birrell, J.M., Brown, V.J., 2000. Medial frontal cortex mediates perceptual attentional set shifting in the rat. Journal of Neuroscience 20, 4320–4324.
Bissonette, G.B., Martins, G.J., Franz, T.M., Harper, E.S., Schoenbaum, G., Powell, E.M., 2008. Double dissociation of the effects of medial and orbital prefrontal cortical lesions on attentional and affective shifts in mice. Journal of Neuroscience 28, 11124–11130.
Bogacz, R., Brown, M.W., Giraud-Carrier, C., 2001. Model of familiarity discrimination in the perirhinal cortex. Journal of Computational Neuroscience 10, 5–23.
Boulougouris, V., Dalley, J.W., Robbins, T.W., 2007. Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. Behavioural Brain Research 179, 219–228.

Boulougouris, V., Robbins, T.W., 2009. Pre-surgical training ameliorates orbitofrontal-mediated impairments in spatial reversal learning. Behavioural Brain Research 197, 469–475.
Bowman, H., Wyble, B., 2007. The simultaneous type, serial token model of temporal attention and working memory. Psychological Review 114, 38–70.
Box, G.E.P., Draper, N.R., 1987. Empirical Model-building and Response Surfaces. Wiley, New York.
Braver, T.S., Barch, D.M., Gray, J.R., Molfese, D.L., Snyder, A., 2001. Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. Cerebral Cortex 11, 825–836.
Bussey, T.J., Saksida, L.M., 2002. The organization of visual object representations: a connectionist model of effects of lesions in perirhinal cortex. European Journal of Neuroscience 15 (2), 355–364.
Butters, N., Butter, C., Rosen, J., Stein, D., 1973. Behavioral effects of sequential and one-stage ablations of orbital prefrontal cortex in the monkey. Experimental Neurology 39, 204–214.
Churchland, P.S., Sejnowski, T.J., 1988. Perspectives on Cognitive Neuroscience. Science 242, 741–745.
Churchland, P.S., Sejnowski, T.J., 1992. The Computational Brain. MIT Press, Cambridge, MA.
Cowell, R.A., Bussey, T.J., Saksida, L.M., 2006. Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex. Journal of Neuroscience 26, 12186–12197.
Cowell, R.A., Bussey, T.J., Saksida, L.M., 2011. Using Computational modeling to understand cognition in the ventral visual-perirhinal pathway. In: Alonso, E., Mondragon, E. (Eds.), Computational Neuroscience for Advancing Artificial Intelligence: Models, Methods and Applications. IGI Global, Hershey, PA, pp. 15–45.
Crick, F., 1989. The recent excitement about neural networks. Nature 337, 129–132.
Dailey, M.N., Cottrell, G.W., 1999. Organization of face and object recognition in modular neural network models. Neural Networks 12, 1053–1074.
Dailey, M.N., Cottrell, G.W., Padgett, C., Adolphs, R., 2002. EMPATH: a neural network that categorizes facial expressions. Journal of Cognitive Neuroscience 14, 1158–1173.
Etcoff, N.L., Magee, J.J., 1992. Categorical perception of facial expressions. Cognition 44, 227–240.
Foutz, T.J., Arlow, R.L., McIntyre, C.C., 2012. Theoretical principles underlying optical stimulation of a channelrhodopsin-2 positive pyramidal neuron. Journal of Neurophysiology 107, 3235–3245.
Fukushima, K., 1980. Neocognition— a self-organizing neural network model for a mechanism of pattern-recognition unaffected by shift in position. Biological Cybernetics 36, 193–202.
Gluck, M.A., Allen, M.T., Myers, C.E., Thompson, R.F., 2001. Cerebellar substrates for error correction in motor conditioning. Neurobiology of Learning and Memory 76, 314–341.
Greve, A., Donaldson, D.I., van Rossum, M.C., 2010. A single-trace dual-process model of episodic memory: a novel computational account of familiarity and recollection. Hippocampus 20, 235–251.
Griffiths, T.L., Chater, N., Kemp, C., Perfors, A., Tenenbaum, J.B., 2010. Probabilistic models of cognition: exploring representations and inductive biases. Trends in Cognitive Sciences 14, 357–364.
Hampshire, A., Chaudhry, A.M., Owen, A.M., Roberts, A.C., 2012. Dissociable roles for lateral orbitofrontal cortex and lateral prefrontal cortex during preference driven reversal learning. Neuroimage 59, 4102–4112.
Hinton, G.E., McClelland, J.L., Rumelhart, D.E., 1986. Distributed representations. In: Rumelhart, D.E., McClelland, J.L. (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1. MIT Press, Cambridge, MA, pp. 77–109.
Huber, D.E., O'Reilly, R.C., 2003. Persistence and accommodation in short-term priming and other perceptual paradigms: temporal segregation through synaptic depression. Cognitive Science 27, 403–430.
Huber, D.E., Shiffrin, R.M., Lyle, K.B., Quach, R., 2002. Mechanisms of source confusion and discounting in short-term priming. 2: Effects of prime similarity and target duration. Journal of Experimental Psychology Learning, Memory, and Cognition 28, 1120–1136.
Huber, D.E., Shiffrin, R.M., Lyle, K.B., Ruys, K.I., 2001. Perception and preference in short-term word priming. Psychological Review 108, 149–182.
Iversen, S.D., Mishkin, M., 1970. Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. Experimental Brain Research 11, 376–386.
Izquierdo, A., Murray, E.A., 2004. Combined unilateral lesions of the amygdala and orbital prefrontal cortex impair affective processing in rhesus monkeys. Journal of Neurophysiology 91, 2023–2039.
Izquierdo, A., Suda, R.K., Murray, E.A., 2004. Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. Journal of Neuroscience 24, 7540–7548.
Jones, A.D., Cho, R.Y., Nystrom, L.E., Cohen, J.D., Braver, T.S., 2002. A computational model of anterior cingulate function in speeded response tasks: effects of frequency, sequence, and conflict. Cognitive, Affective & Behavioral Neuroscience 2, 300–317.
Kilgard, M.P., Merzenich, M.M., 1999. Distributed representation of spectral and temporal information in rat primary auditory cortex. Hearing Research 134, 16–28.
Knight, B.W., 1972. Dynamics of encoding in a population of neurons. Journal of General Physiology 59, 734–766.
Koch, C., 1993. Computational approaches to cognition—the bottom-up view. Current Opinion in Neurobiology 3, 203–208.
Marr, D., 1982. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W.H. Freeman, San Francisco, CA.
McAlonan, K., Brown, V.J., 2003. Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. Behavioural Brain Research 146, 97–103.
McClelland, J.L., Botvinick, M.M., Noelle, D.C., Plaut, D.C., Rogers, T.T., Seidenberg, M.S., Smith, L.B., 2010. Letting structure emerge: connectionist and dynamical systems approaches to cognition. Trends in Cognitive Sciences 14, 348–356.
McClelland, J.L., Mcnaughton, B.L., O'Reilly, R.C., 1995. Why there are complementary learning-systems in the hippocampus and neocortex—insights from the successes and failures of connectionist models of learning and memory. Psychological Review 102, 419–457.
Mcclelland, J.L., Plaut, D.C., 1993. Computational approaches to cognition— top-down approaches. Current Opinion in Neurobiology 3, 209–216.
Meeter, M., Myers, C.E., Gluck, M.A., 2005. Integrating incremental learning and episodic memory models of the hippocampal region. Psychological Review 112, 560–585.
Miller, E.K., Gochin, P.M., Gross, C.G., 1991. Habituation-like decrease in the responses of neurons in inferior temporal cortex of the macaque. Visual Neuroscience 7, 357–362.
Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychological Review 110, 611–646.
Nosofsky, R.M., Kruschke, J.K., Mckinley, S.C., 1992. Combining exemplar-based category representations and connectionist learning rules. Journal of Experimental Psychology-Learning Memory and Cognition 18, 211–233.
O'Reilly, R.C., 1998. Six principles for biologically based computational models of cortical cognition. Trends in Cognitive Sciences 2, 455–462.
O'Reilly, R.C., Munakata, Y., 2000. Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain. MIT Press, Cambridge, MA.
Petersson, M.E., Yoshida, M., Fransen, E.A., 2011. Low-frequency summation of synaptically activated transient receptor potential channel-mediated depolarizations. European Journal of Neuroscience 34, 578–593.
Pissadaki, E.K., Sidiropoulou, K., Reczko, M., Poirazi, P., 2010. Encoding of spatiotemporal input characteristics by a CA1 pyramidal neuron model. PLoS Computational Biology 6, e1001038.
Popper, K.R., 1999. All life is Problem Solving. Routledge, London.
Ratcliff, R., McKoon, G., 2001. A multinomial model for short-term priming in word identification. Psychological Review 108, 835–846.
Raymond, J.E., Shapiro, K.L., Arnell, K.M., 1992. Temporary suppression of visual processing in an RSVP task: an attentional blink? Journal of Experimental Psychology: Human Perception and Performance 18, 849–860.
Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nature Neuroscience 2, 1019–1025.
Rolls, E.T., Deco, G., 2002. Computational Neuroscience of Vision. Oxford University Press, Oxford, England/New York.
Rudebeck, P.H., Murray, E.A., 2008. Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. Journal of Neuroscience 28, 8338–8343.
Schoenbaum, G., Nugent, S.L., Saddoris, M.P., Setlow, B., 2002. Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. Neuroreport 13, 885–890.
Seidenberg, M.S., Mcclelland, J.L., 1989. A distributed, developmental model of word recognition and naming. Psychological Review 96, 523–568.
Sejnowski, T.J., Koch, C., Churchland, P.S., 1988. Computational neuroscience. Science 241, 1299–1306.
Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T., 2007. Robust object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 411–426.
Sohal, V.S., Hasselmo, M.E., 2000. A model for experience-dependent changes in the responses of inferotemporal neurons. Network-Computation in Neural Systems 11, 169–190.
Tait, D.S., Brown, V.J., 2007. Difficulty overcoming learned non-reward during reversal learning in rats with ibotenic acid lesions of orbital prefrontal cortex. Annals of the New York Academy of Sciences 1121, 407–420.
Wallis, G., Rolls, E.T., 1997. Invariant face and object recognition in the visual system. Progress in Neurobiology 51, 167–194.
Zhang, K.C., Ginzburg, I., McNaughton, B.L., Sejnowski, T.J., 1998. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. Journal of Neurophysiology 79, 1017–1044.