UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Advances in techniques for imposing reciprocity in brain-behavior relations

Turner, B.M.; Palestro, J.J.; Miletić, S.; Forstmann, B.U.

DOI
10.1016/j.neubiorev.2019.04.018

Publication date
2019

Document Version
Final published version

Published in
Neuroscience and Biobehavioral Reviews

License
CC BY

Link to publication

Citation for published version (APA):
Turner, B. M., Palestro, J. J., Miletić, S., & Forstmann, B. U. (2019). Advances in techniques for imposing reciprocity in brain-behavior relations. *Neuroscience and Biobehavioral Reviews*, *102*, 327-336. https://doi.org/10.1016/j.neubiorev.2019.04.018

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 Nov 2026

Neuroscience and Biobehavioral Reviews 102 (2019) 327–336

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

Review article

Advances in techniques for imposing reciprocity in brain-behavior relations

Brandon M. Turner[a], James J. Palestro[a], Steven Miletić[b], Birte U. Forstmann[b,*]

[a] Department of Psychology, The Ohio State University, Columbus, OH, USA
[b] Department of Psychology, University of Amsterdam, Amsterdam, Netherlands

ARTICLE INFO

Keywords:
Model-based cognitive neuroscience
Cognitive modeling
Brain mapping

ABSTRACT

To better understand human behavior, the emerging field of model-based cognitive neuroscience seeks to anchor psychological theory to the biological substrate from which behavior originates: the brain. Despite complex dynamics, many researchers in this field have demonstrated that fluctuations in brain activity can be related to fluctuations in components of cognitive models, which instantiate psychological theories. In this review, we discuss a number of approaches for relating brain activity to cognitive models, and expand on a framework for imposing reciprocity in the inference of mental operations from the combination of brain and behavioral data.

1. Introduction

The evolution of technology for measuring brain signals, such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), has provided exciting new opportunities for studying mental processes. Today, scientists interested in studying cognition are faced with many options for relating experimentally-derived neurophysiological variables to the dynamics underlying a cognitive process of interest. While conceptually the presence of these new "modalities" of cognitive measures could have immediately spawned an interesting new integrative discipline, the emergence of such a field has been slow relative to the rapid advancements made in these new technologies. Until a little over a decade ago, much of our understanding of cognition had been advanced by two dominant but virtually non-interacting groups. The largest group, cognitive neuroscientists, relies on models to understand patterns of neural activity brought forth by the new technologies. Like experimental psychologists, the models and methods used by cognitive neuroscientists are typically data-mining techniques, and this approach often disregards the computational mechanisms that might detail a cognitive process. The other group, mathematical psychologists, is strongly motivated by *theoretical* accounts of cognitive processes, and instantiates these theories by developing formal mathematical models of cognition. The models often detail a system of computations and equations intended to characterize the processes assumed to take place in the brain. As a formal test of their theory, mathematical psychologists usually rely on their model's ability to fit and predict behavioral data relative to the model's complexity.

A recent trend in cognitive science is to blend the theoretical and mechanistic accounts provided by models in the field of mathematical psychology with the high-dimensional data brought forth by modern measures of cognition. For example, Forstmann et al. (2011) advocated for the use of *reciprocal* relationships between the latent processes assumed by cognitive models and analyses of brain data. While conceptually, blending these two fields may seem like the ideal approach, as this review will discuss, it is often not straightforward to impose such a relationship (Teller, 1984; Schall, 2004) as there are many theoretical, philosophical, and methodological hurdles any researcher must overcome. Yet, the pursuit continues because the payoff is far too enticing to deter some researchers: the notion that agreed upon theoretical and computational mechanisms supporting cognition could be substantiated in the one organ housing mental operations presents a unique opportunity for major advancements in the understanding of human behavior.

2. Reciprocal relations between brain and behavior

The relationship between fluctuations in neural data and cognitive mechanisms can be assessed through statements about the particular nature of the mapping between neural states and latent cognitive processes (Brindley, 1970; Teller, 1984; Schall, 2004). These mathematical statements are known as *linking propositions*, and they can be formally tested and distinguished. For example, Teller (1984) devised a set of different linking propositions specifying how physiological states map onto psychological states. In Teller's view, linking propositions should be defined by a set of logical relations, and she used systems of relations to define families of linking propositions: identity, similarity, mutual

* Corresponding author.
E-mail addresses: turner.826@gmail.com (B.M. Turner), jpalestro@gmail.com (J.J. Palestro), steven@miletic.nl (S. Miletić), buforstmann@gmail.com (B.U. Forstmann).

https://doi.org/10.1016/j.neubiorev.2019.04.018
Received 30 September 2018; Received in revised form 18 March 2019; Accepted 25 April 2019
Available online 22 May 2019
0149-7634/ © 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

exclusivity, simplicity, and analogy. While these propositions are philosophically desirable, they depend on equality statements, which are impossible to observe in the real world as neurons cannot produce the exact same pattern of firing from one trial to the next. In our view, as trial-to-trial fluctuations in neuronal firing are unlikely to be perfectly predictive of decision dynamics, perfectly axiomatic models can be ruled out. Instead, to practically impose logical relations, we can define statistical relationships that quantify evidence for each logical proposition (see Schall, 2004 for a detailed discussion). Because these statistical relationships are posited to quantify evidence, they are viewed as being mechanically different from perfectly causal models such as those discussed in Pearl (2009), although the intentions may often be similar in spirit. Throughout our review, we will refer to the equations defining statistical relationships as the *linking function*, and will only consider probabilistic links rather than fully causal ones.

The purpose of defining the linking function is to then test which brain areas are related to the psychological variables we care about. In Teller's terms, neurons that form clear logical relationships to psychological states are known as *bridge locus* neurons. In our terms, bridge locus neurons are neurons whose association to psychological variables is quantified through the linking function. In assessing whether brain areas are related to psychological variables, it is vital that we quantify evidence as either confirming or refuting the linking propositions. This way, we will have a clear rule about whether or not brain areas constitute the bridge locus.

Fig. 1 illustrates the concept of the bridge locus, and possible considerations for their instantiation. In each panel, hypothetical brain regions are related to mechanisms within a popular cognitive model, known as the diffusion decision model (DDM; Ratcliff, 1978; Ratcliff and Rouder, 1998; Forstmann et al., 2016). The DDM is useful because it mathematically specifies how psychological variables assumed in the model are related to behavioral variables observed in experiments. For example, consider a choice between detecting leftward and rightward motion in the classic random dot motion task. When viewing the stimulus, we notice small local effects of coherent motion, and over time, we arrive at a general consensus of which of the two motions are more likely. The DDM instantiates this process through sequential sampling: we extract information from the stimulus at each moment in time, and this information is gradually accumulated until we have enough information to make a decision. Conceptually, each response option can be represented in an "evidence" space where the boundary of the evidence space represents the time at which a choice is made. The DDM defines psychological variables in terms of mechanisms, and these mechanisms can be adjusted for individuals or trials to better explain how behavioral data came about. Two of the key mechanisms in the model are the rate of evidence accumulation (i.e., the drift rate illustrated as the black arrow pointing toward a boundary), and the initial evidence for the alternatives (i.e., the starting point of the accumulation process). If we were to relate these mechanisms to brain data (Turner et al., 2015), there are a number of possible linking propositions that should be tested. Considerations in forming the bridge locus are (1) the number of candidate brain regions (one or many), (2) the number of psychological mechanisms (one or many), and (3) which brain regions should be related to which mechanisms in the model.

In the field of model-based cognitive neuroscience, there are now many different approaches for identifying the bridge locus (de Holl et al., 2016; Turner et al., 2017b). Consistent with the mathematical propositions of the bridge locus, several researchers have attempted to infer causality between the two streams of data by either directly replacing mechanisms in cognitive models with neural data, or by searching for brain regions whose statistical properties resemble the statistical properties of cognitive mechanisms. We now review these causally-motivated approaches.

2.1. Direct input

The first approach we consider links neural activity from a given brain area directly to the dynamics of a decision model, and so we refer to it as the direct input approach. One of the issues with using cognitive models such as the DDM is that they are inherently and intentionally abstract. A drift rate defines the rate of accumulation, but what is the drift rate in terms of the neurophysiological process in the brain? Previous research has shown that several areas in the brain, such as the frontal eye field (FEF) and lateral intraparietal (LIP) area, exhibit an "accumulation to threshold" property, where the cumulative sum of their firing rates increases to a threshold level during the decision period, followed immediately by the initiation of a saccade (Bogacz and Gurney, 2007; Boucher et al., 2007; Glimcher, 2003; Hanes and Schall, 1996; Heekeren et al., 2004; Liu and Pleskac, 2011; Mulder et al., 2014a,b; Purcell et al., 2012; Purcell and Palmeri, 2016; Roitman and Shadlen, 2002; Shadlen and Kiani, 2013; Shadlen and Newsome, 2001; Smith and Ratclif, 2004; Summerfield and de Lange, 2013). The pattern exhibited by these neurons is taken to be analogous of the accumulation processes in modern accumulator models, as described above (Brown and Heathcote, 2008; Ratcliff, 1978; Ratcliff and Rouder, 1998; Usher and McClelland, 2001). Because of the striking similarities between the firing of FEF and LIP neurons and the evidence accumulation process in cognitive models, it seems reasonable that the activity in these neurons may map directly onto the accumulation process.

One approach to make accumulator models more concrete is to use the neural activity during a decision process to replace the mathematical mechanism that generates evidence accumulation in the model. This approach tightly constrains the link between neural and behavioral data because the neural data are used to generate a direct prediction about behavioral data in the task. This approach was first explored in Purcell et al. (2010), who mapped the firing rate of neurons in the FEF

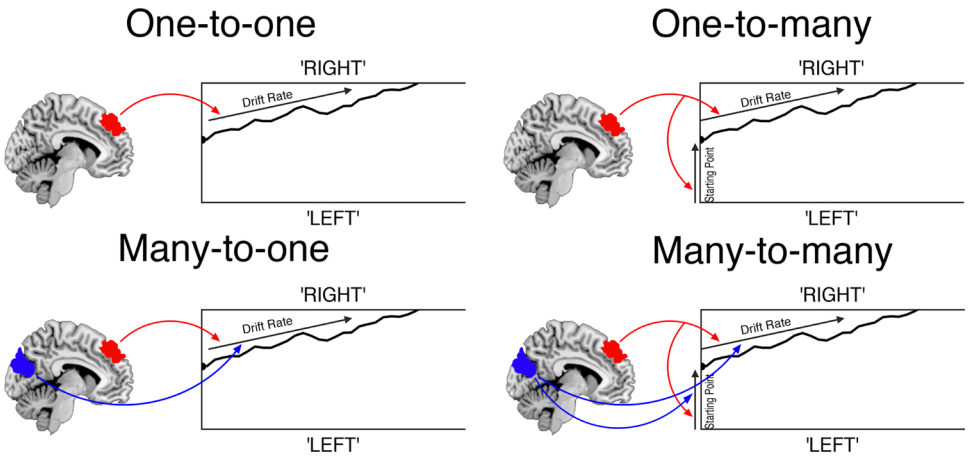


Fig. 1. Considerations when mapping brain to behavior. When forming a map between neural and behavioral data, one must consider the type of connections that must be built, such as the number of brain regions and how they are connected to the mechanisms in a cognitive model. For example, one may consider the joint activation of only a single brain region (top row), a single cognitive mechanism (left column), many brain regions (bottom row), and many cognitive mechanisms (right column).

to the evidence accumulation process in an accumulator model. Specifically, the authors mapped the firing rate of visually responsive neurons within the FEF onto perceptual evidence and the firing rate of movement-related neurons onto evidence accumulation, driving the decision process. Here, the neural activity served as a direct input to the behavioral model, subverting the need for latent processes representing evidence accumulation, such as drift rate and starting point. This allowed for a more explicit test of whether visually responsive and movement-related neurons in ocular motor areas of the brain could predict the onset and location of a saccade in a perceptual decision making task, rather than attempting to explain this process post-hoc by interpreting latent parameter estimates and mapping them onto the proposed underlying mechanisms. Since this initial investigation, several other examples have provided convincing links between the neuronal activity in the ocular motor areas and the dynamics of accumulator models (Purcell et al., 2010, 2012; Purcell and Palmeri, 2016; Schall et al., 2011).

In addition to exploring how the neuronal activity of the FEF could map onto the evidence accumulation process, the direct input approach has been useful in distinguishing between competing accumulator model dynamics, expanding our understanding of how subjects complete the task. By exploiting the constraint imposed between the neural and behavioral data, the authors were able to test the types of transformations of the neural data that were needed to best account for the behavioral data. Specifically, they tested assumptions about how evidence was accumulated over time, such as independent race and counter models (Smith and Van Zandt, 2000; Vickers, 1979), diffusion and random walk models (Laming, 1968; Link and Heath, 1975; Nosofsky and Palmeri, 1997; Ratcliff, 1978; Ratcliff and Rouder, 1998), competing accumulator models (Usher and McClelland, 2001), and gated models (Purcell et al., 2010, 2012; Schall et al., 2011).

Direct input models are most directly related to causal models in the sense that they typically involve direct transformations of the neural signal into decision variables, such as the rate of accumulation in sequential sampling models. Once a transformation has been specified within the model, any fluctuations in the neural data manifest directly as fluctuations in the behavioral response. While in principle this approach has a more concrete link, it places strong assumptions on the veridicality of the neural data, while still treating the behavioral data as probabilistic. This creates some statistical issues when generating the behavioral metrics, as the length of the neural data are explicitly linked to the latency of the behavioral response. For example, Purcell et al. (2010) defined the decision variable as a stochastic process whose primary drive was a direct function of a single unit recording. One can simulate the model using the single unit data up until the length of the neural data has been exceeded. However, if the decision model has not reached a criterion to produce a prediction for the observed behavioral data, how does one go about extrapolating the neural data to continue the stochastic simulation? One approach is to pool information about the single unit data across trials to create an aggregated signal from which simulations can be performed within for example, a condition of the experiment. The pooling approach ensures that a decision can be reached, but it also treats across-trial variability in the neural signal as noise, which distorts the high resolution that single-unit recordings provide. Another solution is to treat the neural data as probabilistic, which in turn creates a statistical rather than purely causal link. Because treating both neural and behavioral data as probabilisitic is more consistent with what we refer to as a joint model, we save the discussion of this alternative approach until later (see Cassey et al., 2016).

2.2. Indirect input

The field of reinforcement learning developed an approach to find neural (often fMRI) correlates of internal model representations (Gläscher and O'Doherty, 2010; O'Doherty et al., 2007). For example, the Rescorla-Wagner model (Rescorla and Wagner, 1972) of classical conditioning characterizes learning as a process of sequentially updating the expected value associated with presented stimuli. The updating process depends on the mismatch between the expected value and actual outcome (the prediction error), modulated by a learning rate parameter that can be estimated using behavioral data. The resulting model produces trial-by-trial expected values and prediction errors, which can be regressed against neural data to find neural correlates of internal representations. Multiple applications of this approach suggest critical roles of ventral striatum in encoding prediction errors and orbitofrontal and mediofrontal cortex in encoding expected value (Daw et al., 2006; Gläscher et al., 2009; Hampton et al., 2006; O'Doherty et al., 2003; Tanaka et al., 2004).

Internal model representations also provide a means to perform model discrimination. Mack et al. (2013) addressed the debate (Minda and Smith, 2002; Zaki et al., 2003) on whether category representations are based on exemplars, or on prototypes (also see Palmeri, 2014). Prototype theory (Posner and Keele, 1968; Reed, 1972) holds that category representations are based on abstract prototypes that bear resemblance to all members of the category, while exemplar theory (Medin and Schaffer, 1978; Nosofsky, 1986) argues that category representations are based on episodic traces formed during learning. Computational models of both theories fit behavioral data equally well, yet the inner representations of both models differ. Mack et al. exploited this discrepancy using multivariate pattern analysis (MVPA) to decode both models' inner representations from fMRI data obtained while participants performed an object categorization task. The results showed that neural data resembled the inner representations of exemplar theory much more closely than those of prototype theory.

Relating internal model representations to neural activity is also a prominent method in the field of vision. For example, the recent success of deep neural networks (DNNs; Kriegeskorte, 2015; LeCun et al., 2015; Yamins and DiCarlo, 2016) in predicting object category spawned research lines to investigate how well DNN representations of visual objects correspond to representations in human cortex (Cadieu et al., 2014; Cichy et al., 2016; Güçlü and van Gerven, 2015; Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014). In one study, Güçlü and van Gerven (2015) transformed DNN representations into predicted neural responses, and correlated these with actual neural responses across the ventral stream of the visual pathway. They showed that the gradient of increasing complexity of object representations across layers in the DNN closely matched the increasing complexity of object representations across the ventral stream. These and similar approaches with other encoding models (Kay et al., 2013a,b; Kay and Yeatman, 2017) help us understand which kind of computations the brain performs to process sensory information into meaningful representations.

2.3. Parametric maps

Where the indirect input approach relates internal model representations to neural data, another approach is to regress the cognitive model parameters themselves onto neural data (Forstmann et al., 2008, 2010a,b, 2016; Boehm et al., 2014; Ho et al., 2012; Mansfield et al., 2011; Mulder et al., 2014, 2012; Summerfield and Koechlin, 2010; van Maanen et al., 2011; White et al., 2014, 2012; Rodriguez et al., 2015; Turner et al., 2018a). Generally, the aim is to explore neural mechanisms underlying cognitive processes. In a now classic example, Forstmann et al. (2008) used this approach to study neural adjustments underlying the speed-accuracy trade-off (SAT) in perceptual decision-making. The SAT refers to the ability to increase accuracy at the cost of speed and vice versa (Bogacz et al., 2010; Heitz and Schall, 2012). Experiments studying the SAT generally instruct participants to stress either speed or accuracy in each upcoming decision-making trial. Decision-making models are then used to quantify the difference in response caution between SAT instructions, and this difference serves as a measure of participants' flexibility in adjusting their behavior.

Forstmann et al. (2008) found that individual variability in response caution adjustments correlate with individual variability in blood oxygenated level dependent (BOLD) responses in striatum and pre-supplementary motor area. Multiple follow-up studies (Mansfield et al., 2011), for example using structural MRI measures (Forstmann et al., 2010b, 2011) or focusing on within-subject variability by calculating trial-by-trial adjustment in response caution (Boehm et al., 2014; van Maanen et al., 2011; Turner et al., 2015), provided additional evidence for a role of these areas in control of response caution. These results are especially interesting as they support prominent theories of action selection in the basal ganglia (Alexander, 1986; Bogacz and Larsen, 2011; Frank, 2006; Lo and Wang, 2006; Ratcliff and Frank, 2012; O'Reilly and Frank, 2006).

Perceptual decision-making models allow researchers to quantify other latent processes as well. Various studies (Forstmann et al., 2010a; Mulder et al., 2012, 2014; Summerfield and Koechlin, 2010) focused on the neural mechanisms that allow for flexible adjustment of behavior due to biasing information. Typically, participants were presented with a cue providing either prior information (i.e., the cued choice option is more likely to be correct), or potential pay-off (i.e., the associated reward with the cued choice option is higher). After quantifying the amount of choice bias using decision-making models, individual differences in bias were correlated with differences in neural measures. The results suggest that in addition to frontoparietal networks (Mulder et al., 2012), the orbitofrontal cortex is involved in processing such bias cues (Forstmann et al., 2010a; Summerfield and Koechlin, 2010). These results imply a role for the orbitofrontal cortex in encoding expected reward, which is corroborated by the reinforcement learning literature described above.

As another example, Turner et al. (2018) examined the relationship between nonlinear mechanisms in decision processes and the engagement of prefrontal cortex in the intertemporal choice task. In this task, subjects are asked to choose between a lower valued immediate reward and a higher valued reward at some point in the future. Similar to the adjustments made in perceptual models for preferential choice (Usher and McClelland, 2004; Hotaling et al., 2010; Turner et al., 2018c), Turner et al. (2018) adapted mechanisms such as lateral inhibition and leakage (intrinsic to Decision Field Theory (Busemeyer and Townsend, 1993) and the Leaky Competing Accumulator model (Usher and McClelland, 2001, 2004); see Box 1 to examine a broad range of possible theoretical explanations of how self-control processes emerge when making goal-directed decisions. Importantly, their analyses revealed that when subjects engage in a self-controlled decision that maximizes reward despite a temporal cost, their brains are differentially activated relative to when they make impulsive decisions that minimize temporal cost and do not maximize reward. After fitting their models hierarchically to data from many individuals, they determined that the best explanation for this neural asymmetry was a dynamic, oscillatory feature selection process (Busemeyer and Townsend, 1993; Hotaling et al., 2010; Dai and Busemeyer, 2014) combined with active asymmetric suppression (i.e., through lateral inhibition; Usher and McClelland, 2001, 2004) of tempting, yet inferior, choice options. Furthermore, Turner et al. (2018) showed how the difference in the asymmetry of active suppression was significantly correlated with fronto-parietal brain areas often engaged in cognitive control (Botvinick et al., 2001, 2004) on a trial-by-trial level.

3. Joint models enforce statistical reciprocity

As discussed in Section 1, linking propositions are strict logical statements between physiological and psychological variables. However, because both neural and behavioral data are noisy and biologically constrained, strict linking propositions are impossible to instantiate formally (Schall, 2004). As a remedy, our methods of assessing the strength of a relationship should be based on statistical principles, where noisy relationships in the data are taken into account. Importantly, to test which brain regions constitute the bridge locus and which do not, we must quantify the strength of the relationship by carefully considering all sources of variability in the neural and behavioral measures. Furthermore, as highlighted in Forstmann et al. (2011), it is important that the link be reciprocal, as both the physiological and psychological bases of the bridge locus are random variables.

One new approach for addressing the statistical uncertainty of the bridge locus while simultaneously imposing a reciprocal link between brain measures and decision variables is the "joint modeling" approach (Turner et al., 2013, 2015, 2016, 2017; Turner, 2015; Cassey et al., 2016). Unlike the direct input or parametric map approaches, joint models enforce a constraint on model parameters based on the random variation in the neural data. In other words, if one treats the neural data as a statistical covariate within the model, the estimates of the behavioral model parameters can be better informed. This simple strategy gives joint models some important advantages. For example, joint models are better equipped to (1) handle mismatching (i.e., when the size of the neural data is different from the size of the behavioral data) and missing data, (2) perform inference on the magnitude of brain-behavior relationships (i.e., they are not subject to Type I errors as in the parametric mapping approach), (3) compare hypothesized brain-behavior relationships across models, and (4) make predictions about either neural or behavioral data.

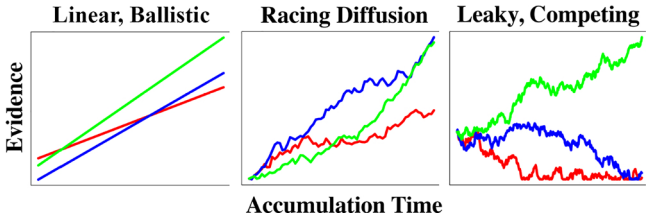
Fig. 2 provides an illustration of the joint modeling approach, applied to 30 s worth of an experiment involving a decision among three alternatives. Neural and behavioral data are separated into streams, and each measure is captured by "submodels." For neural data, candidate sets of ROIs are defined (left panel) and the time course of their activation is extracted. A statistical model of how stimulus presentations (red triangles) affect the BOLD response are specified and fit to the extracted neural signal (middle). The process of fitting the model to data procures estimates of activation parameters δ for each stimulus presentation. For the behavioral data, a cognitive model is developed (left), and similarly fit to behavioral data such as choice response time measures (middle). Parameter estimates θ quantify how the stimulus presentations affect the psychological processes during the task. Finally, to impose statistical reciprocity, a linking function specifies how δ and θ are related (see Box 2).

Of course, there are many different ways of creating a linking function between the two streams of data, and these linking functions have different advantages and disadvantages. One aspect of the linking function that is important for creating divisions in the literature is the manner in which reciprocity is imposed. For example, links can be imposed that are *partially* reciprocal, where only one set of parameters (e.g., δ) are influenced by both streams of data. On the other hand, *fully* reciprocal links can be specified such that both sets of parameters (i.e., θ and δ) are influenced by both streams of data. The definition is a technical one, but it distinguishes the types of reciprocity in terms of how the likelihood function relating model parameters to data is specified. If the likelihood of a stream of data can be written as a function of one (i.e., partial) or both (i.e., full) sets of parameters, it is what we call a joint model. Because the (likelihoods of the) approaches we discuss in Section 1 cannot be expressed in this way, we do not consider them to be joint models per se, although clearly the intentions are similar.

Fig. 3 illustrates three different ways of specifying the linking structure, two of which have been used, and one we will discuss in the Future Directions section below (Fig. 3; right panel). The left panel shows an approach we refer to as a Directed joint model, where neural features are regressed onto model parameters. For example, a linear plane could be used to relate the activation δ_1 and δ_2 of two regions of interest to a model parameter θ (bottom panel). Here, the values of δ_1 and δ_2 strictly determine the value of θ, and so the path of influence is unidirectional, constituting partial reciprocity. Another approach is the Covariance approach, where a probabilistic linking function is imposed.

Box 1
Popular decision models describing human behavior.

There are several models that work well to describe choice response time distributions in a variety of decision making paradigms. Three popular models are the Linear Ballistic Accumulator (LBA; Brown and Heathcote, 2008) model, the Racing Diffusion Model (RDM; Logan et al., 2014), and the Leaky, Competing Accumulator (LCA; Usher and McClelland, 2001) model. These models make a number of different processing assumptions, and the figure below illustrates a few of these important differences. One can view the models as having similar architectures, but with increasing degrees of complexity (arranged in increasing order from left to right).



Linear Ballistic Accumulator Model: The left panel shows a graphical representation of the LBA model for two-choice data. Each response option is represented as a single accumulator (i.e., the red, blue, and green lines). Following the presentation of a stimulus, evidence ballistically accumulates for each alternative until one of the alternatives reaches the threshold (top line). The model assumes some initial amount of evidence is present for each response option, and this amount is randomly distributed across trials. The rate of evidence accumulation itself is also randomly distributed across trials, but has a mean that is fixed allowing one option to be chosen systematically over other options. The accumulation process in the model is linear, and each alternative accumulates information independently, meaning that the state of one accumulator does not depend on any others.

Racing Diffusion Model: The middle panel illustrates a racing diffusion process (Logan et al., 2014), which is a more general case of the DDM. In the racing diffusion process, evidence for each alternative accumulates independently, as in the LBA. However, the DDM assumes that evidence accumulates in a perfectly anti-correlated fashion, meaning that evidence for one alternative is evidence against the other alternative. This feature of the DDM makes it difficult to apply directly to multi-alternative choice. The DDM adds to the LBA an assumption about within-trial variability in the accumulation process. The middle panel illustrates this stochastic process by the wavy paths through evidence space as a function of time.

Leaky Competing Accumulator Model: The LCA model was developed as a neurologically plausible way to describe the dynamics of response competition. Within the LCA, several nonlinearities complicate the accumulation process. Most importantly, the accumulators compete with one another in a way that is state-dependent: as one accumulator gathers more evidence, it can inhibit other accumulators, causing their rate of accumulation to slow and even become negative. In the illustration above, this competitive dynamic can be seen by inspecting the interaction of the accumulators, where the green accumulator dominates first the red accumulator, and later the blue accumulator. Like the DDM, the LCA assumes within-trial variability. Traditional applications of the LCA do not usually assume between-trial variability in the drift rate, and only occasionally assume between-trial variability in starting point. The LCA model also assumes that the accumulation of evidence is "leaky", meaning that some information is lost during the integration of sensory information.

Here, all neural features can interact with one another, as well as the model parameters. The probabilistic map can be used to specify a distribution on θ, where the values of δ_1 and δ_2 are used to slice through a hyper ellipsoid. Here, the path of influence is bidirectional (i.e., double headed arrows), constituting full reciprocity. Finally, to create more flexible maps, one could use a Neural Network approach where all neural features map to "hidden states" before being converted into model parameters. These linking functions allow for distributed activation that can be highly nonlinear, yet still only partial reciprocity is established. We now discuss these linking functions in turn.

3.1. Directed models

The left panel of Fig. 3 illustrates the basic idea behind directed joint models: parameters of a behavioral model θ are linked to parameters of a neural signal of interest δ through a deterministic function.

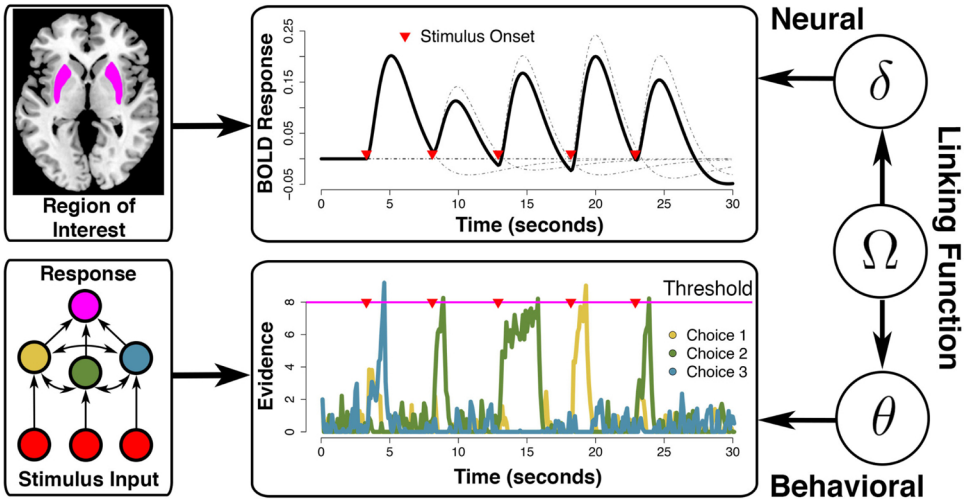


Fig. 2. Illustration of joint modeling approach. The figure shows a hypothetical example consisting of 30 s worth of an experiment involving a decision among three alternatives. For neural data, regions of interest are defined (left) and the blood oxygenated level dependent (BOLD) response can be extracted. Statistical models can be fit to the observed BOLD time course (middle), and parameters δ for say, neural activation, can be estimated. For behavioral data, a cognitive model is developed (left) with mechanisms that are cognitively meaningful. The model can then be fit to data (middle), and parameters θ for say, drift rate, can be estimated. Finally, joint models specify how the neural parameters δ are related to the cognitive model parameters θ through a linking function. In each model schematic, red triangles indicate stimulus presentations.

Box 2
Linking functions relating brain to behavior.

In describing neural data N, one approach is to use a statistical model such as the general linear model (Friston et al., 2002; Penny and Friston, 2004), or topographic latent factor analysis (Gershman et al., 2011). These models have a set of parameters δ that control their shape in ways that can closely match observed neural data. On the other hand, one can use theoretical models of cognitive processes with parameters θ to describe behavioral data B. To complete the process of linking the two streams of data, joint models were proposed as a way to directly relate parameters describing neural data δ to parameters describing behavioral data θ. Turner et al. (2013) proposed a completely generic function of the following form:

$$(\theta, \delta) \sim \mathcal{M}(\Omega). \quad (1)$$

Here, the parameter(s) Ω serve to control the shape of the structure of the link between θ and δ. The connection enforced by the overarching distribution Ω is concrete: one must make a specific assumption about the relationship between θ and δ when considering the underlying cognitive processes involved. As the article has suggested, there are many ways to specify this link, where some links are probabilistic, deterministic, or based on machine learning techniques.

The left panel of Fig. Fig. 3 illustrates the first type of joint model we discussed in this article, an approach we refer to as "Directed" (Cavanagh et al., 2011; Nunez et al., 2015, 2017; Frank et al., 2015). The Directed approach uses a set of parameters δ to describe the functional properties of neural data N through some statistical model that also modulates the behavioral model parameters θ through a linking function $\mathcal{M}$, such that

$$\theta = \mathcal{M}(\delta). \quad (2)$$

In the applications described in this article, the linking function usually takes the form of a multivariate regression model. For example, suppose the parameters $\delta_{i,k}$ describe a set of K activations on Trial i from several regions of interest (i.e., $k \in \{1, 2, ..., K\}$). One could assume a generic linear combination of these activations gives rise to the behavioral parameters of interest, such that

$$\begin{aligned}\theta_i &= \beta_0 + \beta_1\delta_{i,1} + \beta_2\delta_{i,2} + \cdots + \beta_K\delta_{i,K} \\ &= \beta_0 + \sum_k \beta_k\delta_{i,k}.\end{aligned}$$

Here, the activation on each trial for each ROI is scaled by β_k and shifted by β_0 to best capture the neural data, while also generating a good prediction for behavioral data through the parameters θ. This functional form can be viewed as a single-level perceptron model (Minsky and Papert, 1969) that maps a set of inputs δ to a set of outputs θ.

If one cannot assume that there is a direct link between neural and behavioral model parameters, another approach is to specify a probabilistic link between the two parameters. For example, Turner and colleagues (Turner et al., 2013, 2015, 2016; Palestro et al., 2018) have used the multivariate normal distribution to simultaneously model multivariate patterns in neural activation through the form

$$(\theta_i, \delta_{i,1}, \delta_{i,2}, ..., \delta_{i,K}) \sim \mathcal{N}_k(\phi, \Sigma),$$

where ϕ is a set of means for the model parameters, and Σ contains information for the relationship between every pairwise combination of parameters in the set of model parameters. As the complexity of this linking function grows rapidly with increasing number of ROIs (i.e., quadratic complexity), Turner et al. (2017) investigated methods for decomposing Σ into a factor loading matrix Λ, factor variance matrix Φ, and residual terms , such that

$$\Sigma = \Lambda\Phi\Lambda^{\top} + \Psi.$$

This approach has the advantage of constraining the correlation structure on the basis of the model parameters, where the factors within Λ can directly represent parameters from a cognitive model. The factor analytic approach was show to greatly reduce the complexity of the linking function, while preserving the model's out-of-sample generalizability.

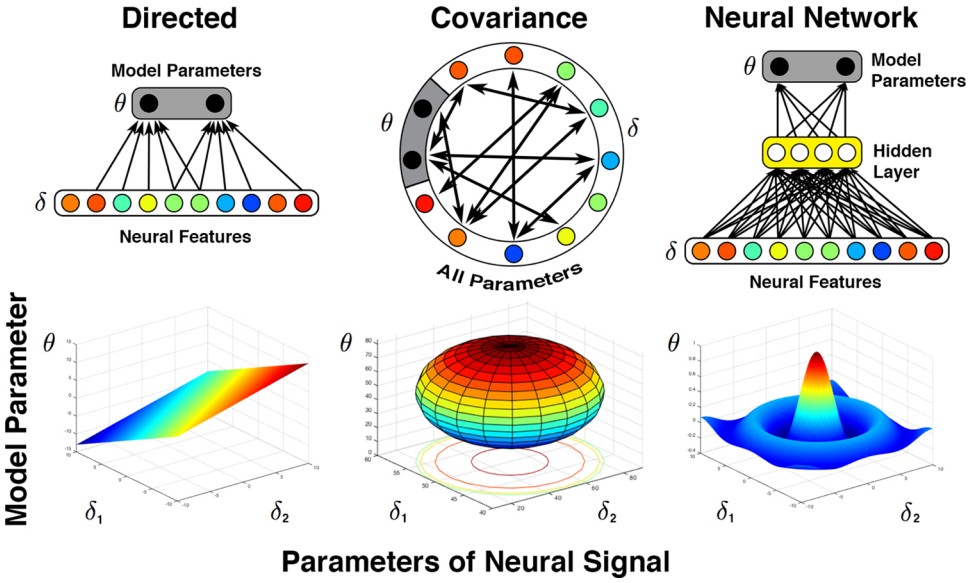


Fig. 3. Different statistical links between brain and theory. There are many ways to specify a linking function between neural and behavioral data (see Box 2). The left panel shows a generic application of a linear regression model. The middle panel shows a multivariate normal linking function that allows variability along each dimension to affect the strength of association. The right panel shows possible new directions for joint models, where multilayer connections between neural data and model parameters can be made to allow for distributed activation and more complex detection of key neural features.

In this way, Directed joint models are quite similar to the direct input and parametric mapping approaches above, but the key difference is that the linking mechanism allows variation in θ to statistically affect variation in δ. Furthermore, consistent with the identification of the bridge locus, we may have different linking structures where several brain regions are related to one or more model parameter. Importantly, the link between θ and δ is reciprocal. Not only do the neural data have a direct effect on the parameters θ, but because the precise form of the linking function makes a strong commitment to a prediction about behavioral data, so too do the behavioral data influence the parameters δ.

At this point, several applications of these directed models have been reported, and they have been particularly effective in perceptual decision making tasks (Cavanagh et al., 2011; Nunez et al., 2015, 2017; Frank et al., 2015; van Ravenzwaaij et al., 2017; Ratcliff et al., 2016; Herz et al., 2017; Hawkins et al., 2017). For example, Nunez et al. (2015) used EEG data on a perceptual decision making experiment as a proxy for attention. They controlled the rate of flickering stimuli presented to subject and measured power of the EEG signal at these frequencies; a measure known as steady-state visual evoked potential. The power on these frequencies is known to be modulated by attention. Importantly, Nunez et al. showed that individual differences in attention or noise suppression was indicative of the choice behavior, specifically it resulted in faster responses with higher accuracy.

In a particularly novel application, Frank et al. (2015) showed how models of reinforcement learning could be fused with the DDM to gain insight into activity in the subthalamic nucleus (STN). In their study, Frank et al. used simultaneous EEG and fMRI measures as covariates in the estimation of single-trial parameters. Specifically, they used predefined regions of interest including the presupplementary motor area, STN, and a general measure of mid-frontal EEG theta power to constrain trial-to-trial fluctuations in response threshold, and BOLD activity in the caudate to constrain trial-to-trial fluctuations in evidence accumulation. Their work is important because it establishes concrete links between STN and pre-SMA communication as a function of varying reward structure, as well as a model that uses fluctuations in decision conflict (as measured by differences in expected rewards) to adjust response threshold from trial-to-trial.

While Fig. 3 illustrates how the parameters δ modulate the parameters θ, other models assume the reverse influence, where the behavioral parameters θ inform the neural parameters δ. As a concrete example, Cassey et al. (2016) extended the single-unit modeling work of Purcell et al. (2010) by linking firing parameters of single unit recordings to evidence accumulation dynamics of a decision model. Cassey et al. modeled data from a seminal experiment by Roitman and Shadlen (2002) containing behavioral recordings of two monkeys in a simple decision-making task. The neural data consisted of single-cell neural recordings from the lateral intraparietal area of the cortex. On each trial, a random dot kinematogram appeared on the screen and the monkey indicated whether the coherently moving dots were drifting left or right. Response times and choices were recorded from each trial as well as the timing of action potentials from a set of neurons in the lateral intraparietal area of the cortex. The joint model builds on the work of Purcell et al. (2010, 2012) by assuming that an evidence accumulation model can provide a tight link between the observed neural firing rate and behavioral responses. In contrast to Purcell et al. (2010, 2012) where the neural data are used directly as input to an evidence accumulation model, the model also included an explicit statistical model of the single unit spike trains. Given this implementation, descriptions of the neural data can be informed by the neuron's properties, such as which neuron was being recorded, and from which monkey. The joint model by Cassey et al. (2016) was hierarchical, and the parameters of the neural submodel were allowed to vary across neurons and monkeys.

3.2. Covariance models

Directed joint models are convenient because of their simplicity, and because they establish a causal role of neural activation in decision making. However, sometimes causal links are too restrictive, and instead what is needed is a probabilistic linking function rather than a deterministic one. For example, the activity in the LIP area has served as the neural basis for the evidence accumulation process (Roitman and Shadlen, 2002; Shadlen and Kiani, 2013; Shadlen and Newsome, 2001), but Katz et al. (2016) have shown that when LIP areas are superficially lessioned in nonhuman primates, patterns of decision making variables remain unaffected. This finding might suggest that while LIP is related to decision variables, they may not be causally linked (Huk et al., 2017).

As Fig. 3 indicates, directed approaches can potentially be too constrained, making the linking structure inflexible for potentially capturing highly complex interactions. As alluded to in Fig. 1, sometimes it will be important to capture multivariate tendencies across several ROIs, or to map brain activity onto multiple model parameters. One way to capture multivariate tendencies is the Covariance approach, which has been used productively to link multiple measures of neural data to pairwise combinations of model parameters (Turner et al., 2013, 2015, 2016, 2017, 2018; Cassey et al., 2016; Palestro et al., 2018). For example, Turner et al. (2013) used structural diffusion weighted imaging data to explain differences in patterns of choice response time data across subjects. They showed how a joint model equipped with information about the interconnectivity of brain areas could make accurate predictions about a subject's behavioral performance in a cross validation test (i.e., the behavioral data were withheld).

Turner et al. (2015) extended this approach to build in brain state fluctuations measured with fMRI into the DDM. The problem Turner et al. (2015) addressed centered on a lack of information about within-trial accumulation dynamics. In behavioral choice response time experiments, following the presentation of a stimulus, researchers can only observe the eventual choice and response time. These data are then used to estimate parameters of a cognitive model, following an assumption that the data observed on each of these trials arises from the same psychological process. However, this assumption – known as stationarity – is a strong one, and is seldom observed in empirical data (Peruggia et al., 2002; Craigmile et al., 2010). Turner et al. (2015) used a multivariate model to describe the joint activation of a set of brain regions of interest, and used this description to enhance the classic DDM. In a cross validation test, they showed that their extended model could generate better predictions about behavioral data than the DDM alone, demonstrating that neurophysiology can be used to improve explanations about trial-to-trial fluctuations in behavior.

In another application, Turner et al. (2016) used the joint modeling framework to perform multimodal data fusion at the individual-subject level. In the study, subjects were assigned to groups that dictated which type of neural measures would be collected: (1) EEG, (2) fMRI, or (3) both EEG and fMRI. Within all groups, subjects completed an intertemporal choice task, providing both behavioral data in the form of choice response times and neural data in accordance with their group assignment. For the subjects providing both EEG and fMRI, Turner et al. used a repeated measures design where subjects returned to the lab and neuroimaging modalities were counterbalanced across individuals. In using the joint model, they assumed that the common relationship of all the measurements (i.e., behavior, EEG, and fMRI) was the mental activity underlying each decision. Using this assumption, Turner et al. created one hierarchical joint model of all three groups, and showed that this model performed better in terms of model fit and cross validation of individual subjects' data compared to models that only considered one (i.e., behavioral only) or two (e.g., behavior and EEG only) modalities of information. Importantly, Turner et al. showed how repeated measures experimental designs can be used to productively integrate information from EEG, fMRI, and behavioral data both within and between individuals.

4. Future extensions: distributed activation

Although Covariance joint models are more flexible than Directed joint models, they still may not provide the best linking function in some scenarios. By capturing all pairwise correlations among ROIs and model mechanisms, they can be computationally complex to fit to data (Turner et al., 2017a). In fact, this complexity has limited Covariance joint models to ROI-based analyses, as whole-brain analyses are currently infeasible. While ROI-based analyses can be a productive way to integrate results across studies, they completely ignore potentially interesting coactivations that may be distributed across the brain (Haxby et al., 2000; Norman et al., 2006). For example, Huk et al. (2017) suggest several reasons why the firing rates of single unit neurons recorded from LIP should not be interpreted as being directly related to decision variables per se, but rather motor preparation signals. Instead, Huk et al. (2017) advocated for the notion that the integration of motion information is distributed across the brain. While such distributed correlation of evidence variables has been observed in ROI-based joint analysis of human fMRI data (Turner et al., 2015, 2017), finer levels of analysis such as whole-brain and temporal dynamics are needed to advance the field.

To capture distributed activations, even more flexible linking functions may be necessary. As noted in Turner et al. (2018), neural networks may provide an interesting opportunity for detecting multivariate coactivation of cortical areas that are not spatially proximal. As illustrated in the right panel of Fig. 3, Neural Network extensions are not unlike Directed approaches; in fact, one can view Directed approaches as a single-layer perceptron model, an early form of connectionist models. Much like the history of the perceptron (Minsky and Papert, 1969; McClelland, 2009), there are likely many types of functions that Directed models are unable to capture. One of the major problems that connectionist frameworks such as PDP models addressed was linearly separable mapping functions such as the XOR operation. In the XOR problem, a map is constructed between two inputs x_1 and x_2 and an output y. When either $x_1 = 1$ or $x_2 = 1$, $y = 1$, but if x_1 ever equals x_2 (i.e., $x_1 = x_2 = 1$ or $x_1 = x_2 = 0$), then $y = 0$. Simple perceptron models are unable to address this type of mapping. The solution to the XOR problem was to include a hidden layer to allow for a more complex mapping function between input and output layers. Analogously, hidden layers may be an essential component to advance Directed joint models for more complex multivariate interactions with cognitive mechanisms. As we see it, Neural Network models, or highly nonlinear multivariate regression techniques in general, serve as a method of agnostically mapping the activation from many neural features into key decision dynamics. There is nothing particularly special about Neural Network models per se, as similar nonlinear multivariate regression techniques could extend Directed joint models to capture similar patterns (see Box 2).

To capture temporal dynamics, one approach would be to model the temporal dependency among regions of interest using techniques such as Dynamic Causal Models (Friston, 2002; Friston et al., 2003), or more generally, Multivariate Dynamical Systems (Ryali et al., 2011). These techniques often require strong assumptions about the set of connectivity paths worth considering, or how activation maps to a hemodynamic signal (e.g., the Balloon model, see Buxton et al., 1998; Mandeville et al., 1999; Friston et al., 2000). Although temporally causal models are generally difficult to fit to data, considerable progress is being made to improve the feasibility of testing these dynamical approaches (Sugihara et al., 2012), and so relating temporal dynamics of brain behavior to the temporal structure of decision models may be the next frontier for imposing reciprocity in brain-behavior dynamics.

5. Conclusions

To connect neuroscientific measures to psychological theory, a new wave of researchers have carefully considered how to inspect and interpret highly complex interactions across a sea of data. Many researchers have looked to computational models that instantiate psychological theories through a set of mathematical expressions, making their predictions for data in completely new experiments transparent. As the field has continued to develop, new statistical techniques have been constructed with the intention of bridging mechanisms from abstract computational models to concrete neurophysiological responses. These powerful new frameworks allow researchers to understand the complexities of brain data in terms of the psychological theories they assume. Some of these frameworks inherently assume hierarchical Bayesian architectures, which have been shown to magnify the resolution of data by the borrowing of “statistical strength.” In closing, techniques such as joint modeling provide the telescope by which neural data may be interpreted through the lens of a cognitive model.

References

Alexander, G., 1986. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu. Rev. Neurosci. 9, 357–381.

Boehm, U., Van Maanen, L., Forstmann, B.U., Van Rijn, H., 2014. Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. NeuroImage 96, 95–105.

Bogacz, R., Gurney, K., 2007. The basal ganglia and cortex implement optimal decision making between alternative actions. Neural Comput. 19, 442–477.

Bogacz, R., Larsen, T., 2011. Integration of reinforcement learning and optimal decision-making theories of the basal ganglia. Neural Comput. 23, 817–851.

Bogacz, R., Wagenmakers, E.J., Forstmann, B.U., Nieuwenhuis, S., 2010. The neural basis of the speed-accuracy tradeoff. Trends Neurosci. 33, 10–16.

Botvinick, M.M., Cohen, J.D., Carter, C.S., 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn. Sci. 8 (12), 539–546.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D., 2001. Conflict monitoring and cognitive control. Psychol. Rev. 108 (3), 624.

Boucher, L., Palmeri, T.J., Logan, G.D., Schall, J.D., 2007. Inhibitory control in mind and brain: an interactive race model of countermanding saccades. Psychol. Rev. 114, 376–397.

Brindley, G.S., 1970. Physiology of Retina and Visual Pathways, second ed. Williams & Wilkins, Baltimore, MD.

Brown, S., Heathcote, A., 2008. The simplest complete model of choice reaction time: linear ballistic accumulation. Cogn. Psychol. 57, 153–178.

Busemeyer, J.R., Townsend, J.T., 1993. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol. Rev. 100, 432–459.

Buxton, R.B., Wong, E.C., Frank, L.R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn. Reson. Med. 39 (6), 855–864.

Cadieu, C.F., Hong, H., Yamins, D.L.K., Pinto, N., Ardila, D., Solomon, E.A., Majal, N.J., DiCarlo, J.J., 2014. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. PLoS Comput. Biol. 10, e1003963.

Cassey, P., Gaut, G., Steyvers, M., Brown, S.D., 2016. A generative joint model for spike trains and saccades during perceptual decision making. Psychon. Bull. Rev. 23, 1757–1778.

Cavanagh, J.F., Wiecki, T.V., Cohen, M.X., Figueroa, C.M., Samanta, J., Sherman, S.J., Frank, M.J., 2011. Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nat. Neurosci. 14, 1462–1467.

Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., Oliva, A., 2016. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Sci. Rep. 6, 1–13.

Craigmile, P., Peruggia, M., Van Zandt, T., 2010. Hierarchical Bayes models for response time data. Psychometrika 75, 613–632.

Dai, J., Busemeyer, J.R., 2014. A probabilistic, dynamic, and attribute-wise model of intertemporal choice. J. Exp. Psychol.: Gen. 143, 1489–1514.

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., Dolan, R.J., 2006. Cortical substrates for exploratory decisions in humans. Nature 441, 876–879.

de Hollander, G., Forstmann, B.U., Brown, S.D., 2016. Different ways of linking behavioral and neural data via computational cognitive models. Cogn. Neurosci. Neuroimaging 1, 101–109.

Forstmann, B.U., Brown, S.D., Dutilh, G., Neumann, J., Wagenmakers, E.-J., 2010a. The neural substrate of prior information in perceptual decision making: a model-based analysis. Front. Hum. Neurosci. 4, 1–12.

Forstmann, B.U., Ratcliff, R., Wagenmakers, E.-J., 2016. Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. Annu. Rev. Psychol. 67, 641–666.

Forstmann, B.U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., Bogacz, R., Turner, R., 2010b. Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. Proc. Natl. Acad. Sci. USA 107, 15916–15920.

Forstmann, B.U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D.Y., Ridderinkhof, K.R., Wagenmakers, E.-J., 2008. Striatum and pre-SMA facilitate decision-making under time pressure. Proc. Natl. Acad. Sci. USA 105, 17538–17542.

Forstmann, B.U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., Brown, S., 2011a. The speed-accuracy tradeoff in the elderly brain: a structural model-based

approach. J. Neurosci. 31, 17242–17249.
Forstmann, B.U., Wagenmakers, E.-J., Eichele, T., Brown, S., Serences, J.T., 2011b. Reciprocal relations between cognitive neuroscience an formal cognitive models: opposites attract? Trends Cogn. Sci. 15, 272–279.
Frank, M.J., 2006. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision-making. Neural Netw. 19, 1120–1136.
Frank, M.J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T.V., Cavanagh, J.F., Badre, D., 2015. fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. J. Neurosci. 35 (2), 485–494.
Friston, K., 2002. Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage 16, 513–530.
Friston, K., Harisson, L., Penny, W., 2003. Dynamic causal modeling. NeuroImage 19, 1273–1302.
Friston, K., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging. NeuroImage 16, 465–483.
Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fmri: the balloon model, volterra kernels, and other hemodynamics. NeuroImage 12 (4), 466–477.
Gershman, S.J., Blei, D.M., Pereira, F., Norman, K.A., 2011. A topographic latent source model for fMRI data. NeuroImage 57, 89–100.
Gläscher, J.P., Hampton, A.N., O'Doherty, J.P., 2009. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb. Cortex 19, 483–495.
Gläscher, J.P., O'Doherty, 2010. Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. WIREs Cogn. Sci. 1, 501–510.
Glimcher, P.W., 2003. The neurobiology of visual-saccadic decision making. Annu. Rev. Neurosci. 26, 133–179.
Güçlü, U., van Gerven, M.A.J., 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. Annu. Rev. Vis. Sci. 1, 417–446.
Hampton, A.N., Bossaerts, P., O'Doherty, J.P., 2006. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J. Neurosci. 8360–8367.
Hanes, D.P., Schall, J.D., 1996. Neural control of voluntary movement initiation. Science 274, 427–430.
Hawkins, G.E., Mittner, M., Forstmann, B.U., Heathcote, A., 2017. On the efficiency of neurally-informed cognitive models to identify latent cognitive states. J. Math. Psychol. 76, 142–155.
Haxby, J.V., Hoffman, E.A., Goblin, M.I., 2000. The distributed human neural system for face perception. Trends Cogn. Sci. 4, 223–233.
Heekeren, H.R., Marrett, S., Bandettini, P.A., Ungerleider, L.G., 2004. A general mechanism for perceptual decision making in the human brain. Nature 431, 859–862.
Heitz, R.P., Schall, J.D., 2012. Neural mechanisms of speed-accuracy tradeoff. Neuron 76, 616–628.
Herz, D.M., Tan, H., Brittain, J.-S., Fischer, P., Cheeran, B., Green, A.L., FitzGerald, J., Aziz, T.Z., Ashkan, K., Little, S., Foltynie, T., Limousin, P., Zrinzo, L., Bogacz, R., Brown, P., 2017. Distinct mechanisms mediate speed-accuracy adjustments in cortico-subthalamic networks. eLife 6, e21481.
Ho, T., Brown, S., van Maanen, L., Forstmann, B.U., Wagenmakers, E.-J., Serences, J.T., 2012. The optimality of sensory processing during the speed-accuracy tradeoff. J. Neurosci. 32, 7992–8003.
Hotaling, J.M., Busemeyer, J.R., Li, J., 2010. Theoretical developments in decision field theory: comment on tsetsos, usher, and chater. Psychol. Rev. 117, 1294–1298.
Huk, A.C., Katz, L.N., Yates, J.L., 2017. The role of the lateral intraparietal area in (the study of) decision making. Annu. Rev. Neurosci. 40, 349–372.
Katz, L.N., Yates, J.L., Pillow, J.W., Huk, A.C., 2016. Dissociated functional significance of decision-related activity in the primate dorsal stream. Nature 535, 285–288.
Kay, K.N., Winawer, J., Mezer, A., Wandell, B.A., 2013a. Compressive spatial summation in human visual cortex. J. Neurophysiol. 110, 481–494.
Kay, K.N., Winawer, J., Rokem, A., Mezer, A., Wandell, B.A., 2013b. A two-stage cascade model of BOLD responses in human visual cortex. PLoS Comput. Biol. 9, e1003079.
Kay, K.N., Yeatman, J.D., 2017. Bottom-up and top-down computations in word- and face-selective cortex. eLife 6, 1–29.
Khaligh-Razavi, S.M., Kriegeskorte, N., 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput. Biol. 10, e1003915.
Kriegeskorte, N., 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. Annu. Rev. Vis. Sci. 1, 417–446.
Laming, D.R., 1968. Information Theory of Choice Reaction Time. Wiley Press, New York, NY.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Link, S.W., Heath, R.A., 1975. A sequential theory of psychological discrimination. Psychometrika 40, 77–105.
Liu, T., Pleskac, T.J., 2011. Neural correlates of evidence accumulation in a perceptual decision task. J. Neurophysiol. 106, 2383–2398.
Lo, C.-C., Wang, X.-J., 2006. Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nat. Neurosci. 9, 956–963.
Logan, G.D., Van Zandt, T., Verbruggen, F., Wagenmakers, E.-J., 2014. On the ability to inhibit thought and action: general and special theories of an act of control. Psychol. Rev. 121, 66–95.
Mack, M.L., Preston, A.R., Love, B.C., 2013. Decoding the brain's algorithm for categorization from its neural implementation. Curr. Biol. 23, 2023–2027.
Mandeville, J.B., Marota, J.J.A., Ayata, C., Zaharchuk, G., Moskowitz, M.A., Rosen, B.R., Weisskoff, R.M., 1999. Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. J. Cereb. Blood Flow Metab. 19 (6), 679–689.
Mansfield, E.L., Karayanidis, F., Jamadar, S., Heathcote, A., Forstmann, B.U., 2011. Adjustments of response threshold during task switching: a model-based functional magnetic resonance imaging study. J. Neurosci. 31 (41), 14688–14692.
McClelland, J.L., 2009. The place of modeling in cognitive science. Top. Cogn. Sci. 1, 11–38.
Medin, D.L., Schaffer, M.M., 1978. Context theory of classification learning. Psychol. Rev. 85, 207–238.
Minda, J.P., Smith, J.D., 2002. Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. J. Exp. Psychol.: Learn. Mem. Cogn. 28, 275–292.
Minsky, M.L., Papert, S.A., 1969. Perceptrons. The MIT Press, Cambridge, MA.
Mulder, M., van Maanen, L., Forstmann, B.U., 2014a. Perceptual decision neurosciences – a model-based review. Neuroscience 277, 872–884.
Mulder, M.J., Boekel, W., Ratcliff, R., Forstmann, B.U., 2014b. Cortico-subthalamic connection predicts individual differences in value-driven choice bias. Brain Struct. Funct. 219, 1239–1249.
Mulder, M.J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., Forstmann, B.U., 2012. Bias in the brain: a diffusion model analysis of prior probability and potential payoff. J. Neurosci. 32, 2335–2343.
Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10, 424–430.
Nosofsky, R.M., Palmeri, T.J., 1997. Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. Percept. Psychophys. 59, 1027–1048.
Nosofsky, R.M., 1986. Attention, similarity, and the identification-categorization relationship. J. Exp. Psychol.: Gen. 115, 39–57.
Nunez, M.D., Srinivasan, R., Vandekerckhove, J., 2015. Individual differences in attention influence perceptual decision making. Front. Psychol. 8 (18), 1–13.
Nunez, M.D., Vandekerckhove, J., Srinivasan, R., 2017. How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. J. Math. Psychol. 76, 117–130.
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 28, 329–337.
O'Doherty, J.P., Hampton, A., Kim, H., 2007. Model-based fMRI and its application to reward learning and decision making. Ann. N. Y. Acad. Sci. 1104, 35–53.
O'Reilly, R.C., Frank, M.J., 2006. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput. 18, 283–328.
Palestro, J.J., Bahg, G., Sederberg, P.B., Lu, Z.-L., Steyvers, M., Turner, B.M., 2018. A tutorial on joint models of neural and behavioral measures of cognition. J. Math. Psychol. 84, 20–48.
Palmeri, T.J., 1997. Exemplar similarity and the development of automaticity. J. Exp. Psychol.: Learn. Mem. Cogn. 23, 324–354.
Palmeri, T.J., 2014. An exemplar of model-based cognitive neuroscience. Trends Cogn. Sci. 18, 67–69.
Pearl, J., 2009. Causality: Models, Reasoning and Inference, second ed. Cambridge University Press, New York, NY.
Penny, W., Friston, K., 2004. Classical and Bayesian inference in fMRI. In: Landini, L. (Ed.), Advanced Image Processing in Magnetic Resonance Imaging. Marcel Dekker, New York.
Peruggia, M., Van Zandt, T., Chen, M., 2002. Was it a car or a cat I saw? An analysis of response times for word recognition. Case Stud. Bayesian Stat. VI, 319–334.
Posner, M.I., Keele, S.W., 1968. On the genesis of abstract ideas. J. Exp. Psychol. 77, 353–362.
Purcell, B.A., Palmeri, T.J., 2016. Relating accumulator model parameters and neural dynamics. J. Math. Psychol. 76, 156–171.
Purcell, B.A., Heitz, R.P., Cohen, J.Y., Schall, J.D., Logan, G.D., Palmeri, T.J., 2010. Neurally-constrained modeling of perceptual decision making. Psychol. Rev. 117, 1113–1143.
Purcell, B.A., Schall, J.D., Logan, G.D., Palmeri, T.J., 2012. Gated stochastic accumulator model of visual search decisions in FEF. J. Neurosci. 32, 3433–3446.
Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85, 59–108.
Ratcliff, R., Frank, M.J., 2012. Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. Neural Comput. 24, 1186–1229.
Ratcliff, R., Rouder, J.N., 1998. Modeling response times for two-choice decisions. Psychol. Sci. 9, 347–356.
Ratcliff, R., Sederberg, P.B., Smith, T.A., Childers, R., 2016. A single trial analysis of EEG in recognition memory: tracking the neural correlates of memory strength. Neuropsychologia 93, 128–141.
Reed, S.K., 1972. Pattern recognition and categorization. Cogn. Psychol. 3, 382–407.
Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A.H., Prokasy, W.F. (Eds.), Classical Conditioning II: Current Research and Theory. Appleton Crofts, New York, pp. 64–99.
Rodriguez, C.A., Turner, B.M., Van Zandt, T., McClure, S.M., 2015. The neural basis of value accumulation in intertemporal choice. Eur. J. Neurosci. 42, 2179–2189.
Roitman, J.D., Shadlen, M.N., 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci. 22, 9475–9489.
Ryali, S., Supekar, K., Chen, T., Menon, V., 2011. Multivariate dynamical systems models for estimating causal interactions in fMRI. NeuroImage 54 (2), 807–823.
Schall, J.D., Purcell, B.A., Heitz, R.P., Logan, G.D., Palmeri, T.J., 2011. Neural mechanisms of saccade target selection: gated accumulator model of the visual-motor cascade. Eur. J. Neurosci. 33, 1991–2002.
Schall, J.D., 2004. On building a bridge between brain and behavior. Annu. Rev. Psychol. 55, 23–50.
Shadlen, M.N., Kiani, R., 2013. Decision making as a window on cognition. Neuron 80, 791–806.

Shadlen, M.N., Newsome, W.T., 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J. Neurophysiol. 86, 1916–1936.
Smith, P.L., Ratclif, R., 2004. Psychology and neurobiology of simple decisions. Trends Neurosci. 27, 161–168.
Smith, P.L., Van Zandt, T., 2000. Time-dependent Poisson counter models of response latency in simple judgment. Br. J. Math. Stat. Psychol. 53, 293–315.
Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. Science 338, 496–500.
Summerfield, C., de Lange, F.P., 2013. Expectation in perceptual decision making: neural and computational. Nat. Rev. Neurosci. 12, 745–756.
Summerfield, C., Koechlin, E., 2010. Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. Front. Hum. Neurosci. 4, 208.
Tanaka, S.C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S., 2004. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat. Neurosci. 7, 887–893.
Teller, D.Y., 1984. Linking propositions. Vis. Res. 24, 1233–1246.
Turner, B.M., 2015. Constraining cognitive abstractions through Bayesian modeling. In: Forstmann, B.U., Wagenmakers, E.J. (Eds.), An Introduction to Model-Based Cognitive Neuroscience. Springer, New York, pp. 199–220.
Turner, B.M., Forstmann, B.U., Love, B.C., Palmeri, T.J., Van Maanen, L., 2017a. Approaches to analysis in model-based cognitive neuroscience. J. Math. Psychol. 76, 65–79.
Turner, B.M., Forstmann, B.U., Steyvers, M., 2018a. Computational approaches to cognition and perception. In: Criss, A.H. (Ed.), Simultaneous Modeling of Neural and Behavioral Data. Springer International Publishing, Switzerland.
Turner, B.M., Forstmann, B.U., Wagenmakers, E.J., Brown, S.D., Sederberg, P.B., Steyvers, M., 2013. A Bayesian framework for simultaneously modeling neural and behavioral data. NeuroImage 72, 193–206.
Turner, B.M., Miletić, S., Forstmann, B.U., 2018b. Outlook on deep neural networks in computational cognitive neuroscience. NeuroImage 180, 117–118.
Turner, B.M., Rodriguez, C.A., Liu, Q., Molloy, M.F., Hoogendijk, M., McClure, S.M., 2018c. On the neural and mechanistic bases of self-control. Cereb. Cortex 29, 732–750.
Turner, B.M., Rodriguez, C.A., Norcia, T., Steyvers, M., McClure, S.M., 2016. Why more is better: a method for simultaneously modeling EEG, fMRI, and behavior. NeuroImage 128, 96–115.
Turner, B.M., Schley, D.R., Muller, C., Tsetsos, K., 2018d. Competing theories of multialternative, multiattribute preferential choice. Psychol. Rev. 125, 329–362.
Turner, B.M., Van Maanen, L., Forstmann, B.U., 2015. Combining cognitive abstractions with neurophysiology: the neural drift diffusion model. Psychol. Rev. 122, 312–336.
Turner, B.M., Wang, T., Merkel, E., 2017b. Factor analysis linking functions for simultaneously modeling neural and behavioral data. NeuroImage 153, 28–48.
Usher, M., McClelland, J.L., 2001. On the time course of perceptual choice: the leaky competing accumulator model. Psychol. Rev. 108, 550–592.
Usher, M., McClelland, J.L., 2004. Loss aversion and inhibition in dynamical models of multialternative choice. Psychol. Rev. 111, 757–769.
van Maanen, L., Brown, S.D., Eichele, T., Wagenmakers, E.-J., Ho, T., Serences, J., 2011. Neural correlates of trial-to-trial fluctuations in response caution. J. Neurosci. 31, 17488–17495.
van Ravenzwaaij, D., Provost, A., Brown, S.D., 2017. A confirmatory approach for integrating neural and behavioral data into a single model. J. Math. Psychol. 76, 131–141.
Vickers, D., 1979. Decision Processes in Visual Perception. Academic Press, New York, NY.
White, C.N., Mumford, J.A., Poldrack, R.A., 2012. Perceptual criteria in the human brain. J. Neurosci. 32, 16716–16724.
White, C.N., Congdon, E., Mumford, J.A., Karlsgodt, K.H., Sabb, F.W., Freimer, N.B., London, E.D., Cannon, T.D., Bilder, R.M., Poldrack, R.A., 2014. Decomposing decision components in the stop-signal task: a model-based approach to individual differences in inhibitory control. J. Cogn. Neurosci. 26, 1601–1614.
Yamins, D.L.K., DiCarlo, J.J., 2016. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 19, 356–365.
Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. USA 111, 8619–8624.
Zaki, S.R., Nosofsky, R.M., Stanton, R.D., Cohen, A.L., 2003. Prototype and exemplar accounts of category learning and attentional allocation: a reassessment. J. Exp. Psychol.: Learn. Mem. Cogn. 29, 1160–1173.