Review

Biological Psychiatry: CNNI

Different Ways of Linking Behavioral and Neural Data via Computational Cognitive Models

Gilles de Hollander, Birte U. Forstmann, and Scott D. Brown

ABSTRACT

Cognitive neuroscientists sometimes apply formal models to investigate how the brain implements cognitive processes. These models describe behavioral data in terms of underlying, latent variables linked to hypothesized cognitive processes. A goal of model-based cognitive neuroscience is to link these variables to brain measurements, which can advance progress in both cognitive and neuroscientific research. However, the details and the philosophical approach for this linking problem can vary greatly. We propose a continuum of approaches that differ in the degree of tight, quantitative, and explicit hypothesizing. We describe this continuum using four points along it, which we dub qualitative structural, qualitative predictive, quantitative predictive, and single model linking approaches. We further illustrate by providing examples from three research fields (decision making, reinforcement learning, and symbolic reasoning) for the different linking approaches.

Keywords: Cognition, Computational models, Functional neuroimaging, Joint modeling, Linking, Mathematical models

http://dx.doi.org/10.1016/j.bpsc.2015.11.004

In recent years, cognitive neuroscientists have applied formal, computational cognitive models to more effectively understand how the brain implements cognitive processes such as decision making, reinforcement learning, and symbolic reasoning. Such formal cognitive models can decompose effects in behavioral data by description in terms of underlying latent cognitive processes and associated variables. Model-based cognitive neuroscience links these variables to brain measurements. This approach can, on the one hand, constrain the development of cognitive models, while, on the other hand, also refine models that explain how cognitive processes are implemented in the brain (1). Linking brain measurements to psychological constructs has been conceptualized as identifying a bridge locus: to link some mental capacity to a neural substrate (2,3). A researcher can identify bridge loci by empirically testing probable linking hypotheses. An example of a linking hypothesis is that the ventral striatum represents how much reward a subject expects during a task.

The scope of this article is limited to the neural linking of computational cognitive models and excludes (much more common) conceptual verbal theories of cognition. A main strength of computational models of cognition over verbal theories is that they force the modeler to be explicit and precise in his or her assumptions about cognition. This reduces the potential for miscommunication and misunderstanding of what a cognitive theory entails and reduces the potential for vague statements that are hard to test empirically (4–6). Additionally, because of their quantitative nature, computational cognitive models offer the possibility of assigning hard numbers to abstract cognitive concepts like response caution or learning rate. These numbers allow the integration of cognitive theory with quantitative neural data in a statistical framework. Ultimately, we believe that this quantitative, statistical approach can bring us much tighter integration between the cognitive and neural domain than verbal theories, supporting more stringent tests of the theories and of the links between neural and behavioral data.

Marr (7) famously subdivided the problem of understanding how the brain works into three levels: 1) a computational level that describes what computational problem a brain aims to solve in a given context, 2) an algorithmic level that describes how the problem can be solved, and 3) an implementational level that describes how this algorithm can physically be performed. Linking cognitive models to neural data can inform theories at all three levels.

For example, at the algorithmic level, cognitive models of speeded decision making make clear predictions about how subjects can lower the distance an evidence accumulator has to travel from the start of a trial to the end. However, for many models, it is not possible to investigate whether this is achieved by increasing the starting point or the finishing threshold of the accumulator with only behavioral data. Clearly, neural data can help to distinguish between these different algorithms and need to carefully be related to the cognitive models that are successful in explaining behavior (5,8).

Similarly, more elaborate explanations at the implementational level are only possible with a firm understanding of what problem the brain is actually solving and what strategies are possible. This point is made again in Marr's original proposal

ISSN: 2451-9022

of the three levels and also eloquently put in the following analogy of David A. Robinson (9): "Trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feather: it just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds' wings make sense." We believe that this calls for the linking of cognitive models, which explain how the computational problems with which the brain is faced can be solved, to neural data, which are rooted in the physical substrate of these algorithms.

Another advantage of linking cognitive models to neural data might be the sheer wealth of additional information that neural data can provide in comparison with behavioral data. By any measure, the amount of information in behavioral data is extremely limited. Because many behavioral experiments provide not much more than choices and reaction times, literally all the data of a behavioral experiment can usually be summarized in a few hundred (only choice) up to a few thousand bytes (also reaction times). Compare that with ultra-high-resolution functional magnetic resonance imaging (fMRI) data from 7 tesla magnetic resonance scanners, which can easily occupy a few billion bytes per subject. Of course, the picture is more complicated than this: the neural data are much more ambiguous. However, recent efforts in sequential sampling models as well as models of value-based learning have taught us that to reliably estimate the parameters of more complicated cognitive models and dissociate between different versions of them, the amount of information of most behavioral datasets is very limited (10,11). Thus, even disregarding the conceptual benefits, cognitive modelers should welcome the practical benefits arising from the wealth of extra information in neural data, as they provide an opportunity to develop richer models of cognition than has been possible so far.

But how do we link cognitive models to functional brain measurements most effectively? In the past decade, parameters of formal cognitive models have been linked to many measures of neural activity, such as electroencephalography (EEG), fMRI, and single-cell recordings. These studies employed wildly varying approaches, connecting variability in behavior and brain measurement at the level of subjects, conditions, and even trials. In some studies, cognitive models were used to set up testable hypotheses about brain activity. In other studies, cognitive model parameters were directly correlated against measurement models of neural data, after both models were fit to their respective data domain. Some studies made a single model of both brain and behavior and tried to predict both at the same time.

In this review, we aim to provide a particular taxonomy of possible methods of linking neural data to cognitive models. We think this taxonomy is useful to describe the work that has been done so far and understand how it has progressed. Additionally, it offers cognitive neuroscientists a set of handles on where to start when linking neural data to cognitive models, as well as what to strive for in the long run (see also the Discussion).

We then give some examples of the four categories of linking in three subfields of cognitive neuroscience from the literature. A larger review of the literature can be viewed in Supplement 1.

Finally, we will discuss the strengths and weaknesses of different points on the continuum and lay out future challenges and developments.

LOOSER AND TIGHTER LINKS

There are many approaches to linking formal models of cognition to neural data. These approaches differ in how explicit and precise the link is made between neural, physiological processes on the one hand and cognitive, phenomenal processes on the other hand. We propose a continuum of tightness of linking. At the loosest level, cognitive models can be linked with neural data simply by constraining the kinds of structural assumptions allowed in the models in order to respect data about neural structures. Tighter links can be created by comparison of predictions for neural and behavioral data or neural and behavioral model parameters. The very tightest and most explicit links are specified by joint models, which make quantitative predictions about both neural and behavioral data at the same time.

Table 1 provides some illustrative examples that are elaborated below. These examples highlight four commonly used points on the continuum between loose to tight linking. We first provide definitions for the four different commonly used levels of linking. Following that, we give detailed examples of these approaches in practice, with each level of linking illustrated in up to three different research domains: perceptual decision making, reinforcement learning, and symbolic reasoning.

Qualitative Structural Linking

Neural data on the structure of the brain are used to constrain the structure of a cognitive model. An example of this is the leaky competing accumulator (LCA) model: "The principles included in the modeling effort have neurobiological as well as computational or psychological motivation, and the specific instantiations of the principles are informed by additional neurophysiological observations" (12).

Qualitative Predictive Linking

A cognitive model is tested using qualitative predictions about both neural and behavioral data. For example, Borst *et al.* (13) used the symbolic reasoning modeling framework of Adaptive Control of Thought-Rational (ACT-R) to make predictions about the difference in fMRI signals between conditions that differed in behavioral measures associated with task difficulty separately for different brain regions: "The model does not predict a general increase in BOLD [blood oxygen-level dependent] response with task difficulty; instead, it predicts lower but more persistent activation levels for the more difficult conditions in the visual and manual modules, and higher and more persistent activation levels for the more difficult conditions in the problem state and declarative memory modules" (13).