ARTICLE IN PRESS

Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Integrating theoretical models with functional neuroimaging

Michael S. Pratte[a,b,*], Frank Tong[b]

[a] Department of Psychology, Mississippi State University, United States

[b] Department of Psychology and the Vanderbilt Vision Research Center, Vanderbilt University, United States

ARTICLE INFO

Article history:
Available online xxxx

ABSTRACT

The development of mathematical models to characterize perceptual and cognitive processes dates back almost to the inception of the field of psychology. Since the 1990s, human functional neuroimaging has provided for rapid empirical and theoretical advances across a variety of domains in cognitive neuroscience. In more recent work, formal modeling and neuroimaging approaches are being successfully combined, often producing models with a level of specificity and rigor that would not have been possible by studying behavior alone. In this review, we highlight examples of recent studies that utilize this combined approach to provide novel insights into the mechanisms underlying human cognition. The studies described here span domains of perception, attention, memory, categorization, and cognitive control, employing a variety of analytic and model-inspired approaches. Across these diverse studies, a common theme is that individually tailored, creative solutions are often needed to establish compelling links between multi-parameter models and complex sets of neural data. We conclude that future developments in model-based cognitive neuroscience will have great potential to advance our theoretical understanding and ability to model both low-level and high-level cognitive processes.

1. Introduction

The general goal of cognitive psychology has been to understand psychological processes at what Marr (1982) would call the *algorithmic* or *representational level* (see Love, 2015). In order to explore the algorithms and representational structures that might underlie processes such as attention or memory, cognitive psychologists often propose formal theoretical models of these processes, and test them by assessing predicted patterns in behavior. For decades, doing so has provided remarkable insights into how the mind works. During the same time, the field of neuroscience has made strides in understanding what Marr calls the *implementation* level, or, how these processes are implemented in the biological machinery that makes up the brain. More recently, the merger of these fields into a unified *cognitive neuroscience* has resulted in part from the development of new neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), which have made investigating the biological substrates of human cognition possible. The more targeted approach of combining theoretical modeling and neuroscience has been termed *computational neuroscience*, a field that is often credited as originating from Marr's work. In this review, we consider a newly emerging endeavor that has arisen from the merger of cognitive psychology, theoretical modeling and neuroscience: using theoretical models in conjunction with human neuroimaging to study psychological processes.

Advances in mathematical and computational approaches have played a key role in fMRI since its invention. Some of these developments include analytical approaches for extracting relevant information from the BOLD signal across the temporal domain, such as the use of temporal phase-encoded designs (Engel, 2012; Engel et al., 1994; Sereno et al., 1995) and the development of de-convolution approaches for fast-event related designs (Boynton, Engel, Glover, & Heeger, 1996; Buckner et al., 1996; Glover, 1999). Developments in inferential statistical techniques have produced a number of tools that have helped make fMRI mapping studies so successful, especially regarding techniques to account for what is arguably the most serious multiple comparisons problem in psychology (see Nichols & Hayasaka, 2003). Given the multivariate nature of fMRI data, correlation-based approaches (Haxby et al., 2001), machine learning techniques (Kamitani & Tong, 2005; Norman, Polyn, Detre, & Haxby, 2006; Tong & Pratte, 2012) and voxel-based modeling approaches (Brouwer & Heeger, 2009; Kay, Naselaris, Prenger, & Gallant, 2008; Serences & Saproo, 2009) have been used to capture the complexities of these high-dimensional data sets, providing powerful new ways of identifying perceptual

* Correspondence to: Mississippi State University, Department of Psychology, 255 Lee Blvd, Starkville, MS 39762, United States.
E-mail address: prattems@gmail.com (M.S. Pratte).

http://dx.doi.org/10.1016/j.jmp.2016.06.008
0022-2496/

information contained in brain activity patterns (Harrison & Tong, 2009; Haynes & Rees, 2005; Kamitani & Tong, 2005; Serences & Boynton, 2007), as well as evidence of semantic tuning properties (Huth, Nishimoto, Vu, & Gallant, 2012; Mitchell et al., 2008). The correlational structure of activity patterns has also been used to characterize object representations in the ventral temporal lobe (Haxby et al., 2001; Kriegeskorte et al., 2008). Correlation-based approaches have also proven useful for delineating the functional connectivity of the human brain (e.g., Honey et al., 2009), including resting state networks (Fox et al., 2005), and more recent studies of brain connectivity have benefitted from the application of graph theoretical models (Bullmore & Sporns, 2009) and other model-based approaches (Tavor et al., 2016). Such advances in analytic methods continue to expand the ways in which fMRI can be used to study the brain.

More recently, research on new mathematical approaches for fMRI has evolved beyond the goal of simply developing more powerful analytic methods, to that of integrating and testing cognitive models. This model-based approach to cognitive neuroscience represents an exciting development that goes beyond the simpler goals of "brain mapping", identifying correlations between individual differences and brain activity, or information-based approaches to characterize cortical function. Instead, the goal of model-based cognitive neuroscience lies in describing the perceptual or cognitive processes that underlie behavior in a mathematically precise manner, and determining the neural processes that underlie these computations.

Cognitive process models have a long history in the study of human performance. For example, models based on signal detection theory have served as the foundation for studying perception (Green & Swets, 1966), attention (Lu & Dosher, 1998) and memory (Kintsch, 1967). Early cognitive research also demonstrated that stochastic accumulator models can accurately predict patterns of choice reaction times across numerous behavioral paradigms (Ratcliff & Rouder, 1998; Stone, 1960). While some cognitive process models have focused on identifying and quantifying a few key parameters to capture patterns of cognitive performance, other models rely on general learning principles to train complex networks with numerous parameters to perform a cognitive task. For example, neural network models (e.g., McClelland & Rogers, 2003) have been developed to characterize high-level processes including speech perception (McClelland & Elman, 1986), categorization (Ashby & Maddox, 1993; Nosofsky & Palmeri, 1997), cognitive control (Botvinick, Braver, Barch, Carter, & Cohen, 2001), and human memory (Polyn, Norman, & Kahana, 2009).

One might expect that the application of theoretical models to neuroimaging data would be a naturally obvious and fruitful endeavor. However, most cognitive neuroscientists have not rushed to meet this challenge until recently. Why has this been the case? A central challenge lies in establishing strong links between the parameters of a cognitive model and particular brain responses embedded within a cognitive experiment. Cognitive models typically rely on latent constructs of presumed psychological processes that must somehow be translated into a predicted pattern of brain responses. If the model leads to clear predictions regarding how the univariate BOLD responses should change over time, such models may be more straightforward to test using standard fMRI analysis procedures. Earlier applications of model-based fMRI have relied on such approaches to identify the neural correlates of reward prediction error (e.g., O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; O'Doherty et al., 2004; O'Doherty, Hampton, & Kim, 2007) and response conflict (e.g., Botvinick, Cohen, & Carter, 2004). However, fMRI data is very high-dimensional, such that establishing links between cognitive models and the information contained in multivariate brain activity patterns is considerably more challenging. Even if a model can be positively related to an information-based metric of brain processing, the next step of determining whether a particular model provides a compelling fit of the high-dimensional brain data can be difficult to demonstrate.

In this review, we highlight several recent studies that have successfully combined theoretical models with fMRI data, addressing diverse questions spanning lower-level perceptual processes to higher-level cognitive processes. A central theme across these studies is the goal of identifying compelling relationships between brain, model and behavior (see Fig. 1(A)), often with the cognitive model serving as the intermediary for mapping between brain and behavior. However, as we will see, there are many possible options and approaches for establishing these links, as a model fitted to behavioral data might be used to predict brain responses, or brain data might be incorporated into a model to predict behavior. Moreover, intermediate processing steps may take place before links are established, such as methods to reduce the high dimensionality of brain data to lower-dimensional measures that can be more directly related to model predictions.

We begin this review by discussing an application of the normalization model to the visual perception of orientation (Brouwer & Heeger, 2011). In this work, fMRI data from early cortical visual areas was first transformed into interpretable constructs using a multivariate modeling approach, and the normalization model was then fitted to the resulting measurements. We then describe an application of models of visual attention to fMRI data (Pratte, Ling, Swisher, & Tong, 2013). Here, a multivoxel pattern classification approach was used to transform multivariate fMRI data, obtained from multiple levels of the visual hierarchy, into an interpretable measure of information representation, and the theoretical model was fitted to the result. Whereas in both of these studies, a formal model was fitted to information decoded from the multivariate fMRI signal (Fig. 1(B)), we next consider a study in which the time course of the fMRI signal on individual trials was incorporated within a theoretical model of behavioral memory performance (Fig. 1(C)), to determine whether this neural signal can lead to more accurate predictions of free recall performance (Kragel, Morton, & Polyn, 2015). An fMRI study of categorization highlights yet another approach (Fig. 1(D)), by assessing the degree to which competing models of behavioral categorization performance can account for observed patterns of neural data (Mack, Preston, & Love, 2013). Finally, we review a study of cognitive control that demonstrates how the application of a theoretical model to fMRI data can reveal new insights about neural processing that would have been impossible without the model (Ide, Shenoy, Yu, & Li, 2013). Here, a model of behavioral performance in the stop-signal task was incorporated within the fMRI analysis (Fig. 1(E)), and the results suggest that the function of the anterior cingulate is more specific than has been suggested by previous studies.

This collection of studies demonstrates both the feasibility and potential of model-based cognitive neuroscience. The approaches are remarkably diverse, both in how the data are used to inform the model, and in the technical solutions employed to establish compelling relationships between brain, model and cognitive performance (see Fig. 1). Of particular interest is that fact that none of the reviewed works share exactly the same strategy to link a theoretical model with functional imaging data. Rather, these examples underscore how individual studies have relied on clever innovations that are custom-built for a particular model or experimental paradigm. As such, we neither foresee nor prescribe a one-size-fits-all approach to model-based cognitive neuroscience. Instead, we believe that the diversity in attempts to integrate modeling and functional brain imaging will forward the advance of theoretical models with a momentum that would not happen if these fields remained isolated from one another.

Please cite this article in press as: Pratte, M. S., & Tong, F., Integrating theoretical models with functional neuroimaging. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.06.008