Neurobiology of Learning and Memory 109 (2014) 193–206

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

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

Invited Review

Combining fMRI and behavioral measures to examine the process of human learning

Elisabeth A. Karuza *,[1], Lauren L. Emberson *,[1], Richard N. Aslin

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA

A R T I C L E I N F O

Article history:
Received 17 April 2013
Revised 16 September 2013
Accepted 18 September 2013
Available online 25 September 2013

Keywords:
fMRI
Cognitive neuroscience
Learning
Process
Time-course
Outcome measures
Behavioral measures

A B S T R A C T

Prior to the advent of fMRI, the primary means of examining the mechanisms underlying learning were restricted to studying human behavior and non-human neural systems. However, recent advances in neuroimaging technology have enabled the concurrent study of human behavior and neural activity. We propose that the integration of behavioral response with brain activity provides a powerful method of investigating the process through which internal representations are formed or changed. Nevertheless, a review of the literature reveals that many fMRI studies of learning either (1) focus on outcome rather than process or (2) are built on the untested assumption that learning unfolds uniformly over time. We discuss here various challenges faced by the field and highlight studies that have begun to address them. In doing so, we aim to encourage more research that examines the process of learning by considering the interrelation of behavioral measures and fMRI recording during learning.

1. Introduction

Relatively recent advances in neuroimaging technology, specifically functional magnetic resonance imaging (fMRI), have made possible the large-scale study of neural systems underlying learning in the human brain. Learning, or the experience-based process by which we form representations of the world around us, is a topic particularly suited to investigation using neuroimaging. Crucially, fMRI has the potential to reveal fluctuations in neural activity as learning unfolds over time, thereby allowing researchers to tap into the *process* through which internal representations undergo change.

Despite the natural fit between fMRI and the study of learning, a critical review of the relevant literature reveals that studies tend to address the question of which brain areas subserve retrieval or recognition of *already learned* items, not the process which generates changes in representation in the first place. While an outcome-focused approach is certainly valuable and germane to the study of learning, we suggest that the time is ripe for the field to focus instead on the *process* of acquisition rather than its outcomes. In this review, we will discuss some corresponding challenges, both methodological and theoretical in nature, and offer suggestions for improved experimental design and analysis. Specifically, we will consider the benefits of incorporating on-line behavioral testing and the use of computational models to predict the time-course of learning. Along these lines, we will discuss those studies that have begun to surmount these challenges to begin to uncover the neural systems engaged over the time-course of learning. This review also explores the critical yet open question of how to interpret the neural changes that occur before behavioral evidence of learning emerges. We begin by considering the unique niche of fMRI in the study of learning and the ways in which this methodology has already shaped and been shaped by the field of cognitive neuroscience.

1.1. What fMRI has already offered the study of learning

For over a century, there has been a voluminous and fruitful tradition of research aimed at the general study of learning in both animals and humans (Rescorla, 1988; Shanks & St. John, 1994; Skinner, 1938; Thorndike, 1931; Tolman, 1951). Until very recently, most of our understanding of this process has relied on studies of either human behavior (e.g., through learning tasks and behavioral manipulations) or non-human neural systems (e.g., through electrophysiological recordings). While research on the neural mechanisms of human learning has benefitted from examining the effects of brain lesions, this case study approach has limited power in revealing the neural systems supporting cognitive mechanisms (Zurif, Swinney, & Fodor, 1991; but see also Caramazza & Badecker, 1991).

* Corresponding authors. Address: Department of Brain & Cognitive Sciences, Meliora Hall, RC 270268, University of Rochester, Rochester, NY 14627, USA.

E-mail addresses: ekaruza@bcs.rochester.edu (E.A. Karuza), lemberson@bcs.rochester.edu (L.L. Emberson).

[1] These authors contributed equally to this manuscript.

1074-7427/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.nlm.2013.09.012

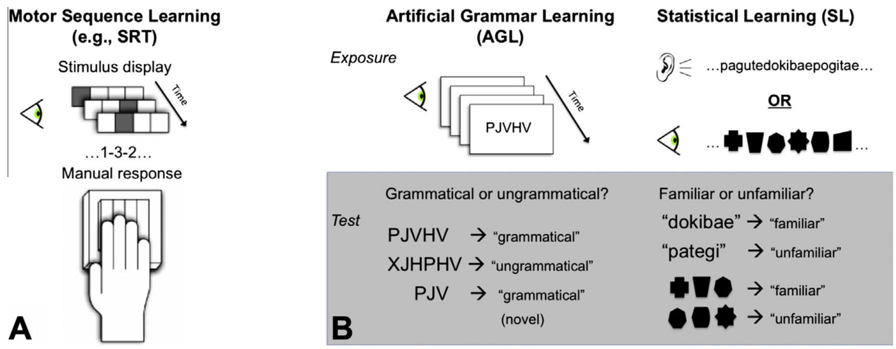

Fig. 1. Structure of various learning paradigms relevant to this review. (A) *Motor Sequence Learning*: Participants initiate a motor response to temporally-patterned visual stimuli. (B) *Artificial Grammar Learning* and *Statistical Learning*: Participants undergo an exposure phase during which they are presented with finite-state grammar sequences (AGL) or probabilistic auditory/visual patterns (SL). In a subsequent test phase, they make acceptability judgments on structured and unstructured test items.

While our review certainly intersects with the study of learning in general, we will focus on the interrelated and overlapping research areas of (1) incidental learning, or acquisition in the absence of specific intention to learn; (2) statistical learning, or acquisition of structural representations via distributional regularities in sensory input; and (3) sequence learning, or acquisition of sequential information across perceptual modalities using both motor measures (motor sequence learning) and measures not specifically focused on motor responses.[2]

In the past decade or so, the emerging use of neuroimaging methods, particularly fMRI, has provided the unique opportunity to study the relationship between neural systems and behavior in human participants on a large scale. While other neuroimaging modalities such as electroencephalography (EEG) and positron emission tomography (PET) have been employed to study learning, fMRI offers a combination of unambiguous spatial location of signals (as opposed to EEG), while being less invasive and safer than PET with high functional and anatomical image resolution (not available using either EEG or PET). The significant technological advances of this method have resulted in a veritable explosion in fMRI studies of human learning.

In addition to opening up the door to the concurrent study of behavior and neural activity in humans, fMRI allows for the (virtually) simultaneous recording of activity across the entire brain and, for some tasks, across the entire time-course of learning. Indeed, the use of fMRI has already enabled investigations into learning that had been tortuous or impossible when using historically employed methods. For example, the theory that the basal ganglia and hippocampus comprise multiple, dissociable learning and memory systems has been investigated and supported using lesion studies in both human and non-human animals and electrophysiology (see Eichenbaum and Cohen (2001) for an excellent historical review). However, the ability to record activity across the entire human brain with fMRI has enabled the *in vivo* investigation of the activity of both of these systems, allowing researchers to ask questions such as, are the basal ganglia and hippocampus simultaneously active during a single learning task (Poldrack, Prabhakaran, Seger, & Gabrieli, 1999)? If not, then do they directly inhibit each other (Poldrack et al., 2001)? Are there some tasks where these systems complement each other (Shohamy & Wagner, 2008)? While the answers to such questions remain elusive, this example serves to illustrate the way in which the goal-directed use of fMRI has fueled productive discussion and advanced our understanding of learning. Thus, fMRI has already provided new avenues to consider the interrelationship between functional neural activity and human learning.

While fMRI continues to be a popular and powerful method for answering a variety of empirical questions, no single method can fully delineate a system as complex as the human brain. FMRI is no exception, in part because it is an indirect measure of the neural activity in the brain that results from changes in blood oxygenation (the blood-oxygen-level-dependent or BOLD response). It has been well established that the BOLD response can be stimulus-evoked (e.g., Belliveau et al., 1991; Ogawa et al., 1992) and, by extension, sensitive to functional neural activity. However, the specific aspects of the neural signal producing the BOLD response are still not entirely clear. Logothetis and Wandell (2004) propose that the BOLD response best corresponds to local field potentials (LFPs) rather than spiking activity directly. Of course, these two aspects of the neural signal are interrelated, but LFPs and spiking pick up on separable aspects of the neural signal; LFPs reflect sub-threshold integrative processes or computations on the input of neural signals, while spiking reflects the output of this computation. If the BOLD response does reflect LFPs more directly than spiking, fMRI can then be considered complementary to the spiking activity typically gathered using electrophysiological methods. It therefore follows that fMRI and electrophysiology can be seen as distinct but highly compatible methods capable of probing neural computations.

1.2. The current use of FMRI to study learning

Given the current impact and future potential of fMRI as a method of investigating human learning, it is perhaps surprising that a significant number of fMRI studies dealing with this topic have elected to focus on the outcome of learning (Forkstam, Hagoort, Fernandez, Ingvar, & Petersson, 2006; Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004; Petersson, Folia, & Hagoort, 2012; Petersson, Forkstam, & Ingvar, 2004; Seger, Prabhakaran, Poldrack, & Gabrieli, 2000; Skosnik et al., 2002; Yang & Li, 2012). That is, fMRI recordings are typically collected during tests of already acquired information, not during the initial processing of structured stimuli (henceforth referred to as the *exposure/acquisition* phase). In such cases, the extent of learning is often measured using post-acquisition tasks that typically involve novelty detection and/or accuracy judgments. Implicit in this approach is that learning is a relatively uniform, time-invariant process that can be examined using post-acquisition outcome measures. Importantly, it has been largely unstudied whether the neural

[2] The majority of studies falling into one or more of these three categories purportedly involve implicit learning or learning without conscious awareness. While the distinction between explicit and implicit forms of learning (both neurally and behaviorally) remains a major area of active debate, we elect not to make any strong claim as to the extent to which the learning studies covered here are wholly implicit or wholly explicit. In order to focus on the broader points laid out above, we will neither weigh in on this debate nor discuss any differences between the studies reviewed along this dimension.