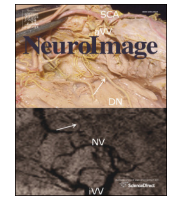
NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

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



Physics instruction induces changes in neural knowledge representation during successive stages of learning

Robert A. Mason *, Marcel Adam Just

Center for Cognitive Brain Imaging, Psychology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA

A R T I C L E I N F O

Article history:
Accepted 29 December 2014
Available online xxxx

Keywords:
Learning
fMRI
Causality
Mechanical systems
Knowledge representation

A B S T R A C T

Incremental instruction on the workings of a set of mechanical systems induced a progression of changes in the neural representations of the systems. The neural representations of four mechanical systems were assessed before, during, and after three phases of incremental instruction (which first provided information about the system components, then provided partial causal information, and finally provided full functional information). In 14 participants, the neural representations of four systems (a bathroom scale, a fire extinguisher, an automobile braking system, and a trumpet) were assessed using three recently developed techniques: (1) machine learning and classification of multi-voxel patterns; (2) localization of consistently responding voxels; and (3) representational similarity analysis (RSA). The neural representations of the systems progressed through four stages, or states, involving spatially and temporally distinct multi-voxel patterns: (1) initially, the representation was primarily visual (occipital cortex); (2) it subsequently included a large parietal component; (3) it eventually became cortically diverse (frontal, parietal, temporal, and medial frontal regions); and (4) at the end, it demonstrated a strong frontal cortex weighting (frontal and motor regions). At each stage of knowledge, it was possible for a classifier to identify which one of four mechanical systems a participant was thinking about, based on their brain activation patterns. The progression of representational states was suggestive of progressive stages of learning: (1) encoding information from the display; (2) mental animation, possibly involving imagining the components moving; (3) generating causal hypotheses associated with mental animation; and finally (4) determining how a person (probably oneself) would interact with the system. This interpretation yields an initial, cortically-grounded, theory of learning of physical systems that potentially can be related to cognitive learning theories by suggesting links between cortical representations, stages of learning, and the understanding of simple systems.

Introduction

Just knowing what some mechanical system accomplishes is often sufficient. Sometimes it is enough to know that when you aim a fire extinguisher and squeeze the handle, a fire-suppressing fluid sprays onto the fire. But how does it work? Learning how a mechanical device works through instruction is a critical part of many jobs. Understanding the psychological and neural processes that occur during such learning can now be studied with brain imaging to reveal how new technical knowledge is built up in the brain in the course of instruction.

A main aim of this research was to show how the neural representations of specific technical knowledge change as a result of acquiring new information. Of course, there are many other types of changes in the brain following training or instruction that have been reported. There are three critical differences between prior investigation of brain changes due to learning and this research, measured here are: (1) changes in representation as opposed to structural changes; (2) changes due to instruction-based learning as opposed to training, specifically in the science domain of the physics of mechanical systems; and (3) changes in the neural representation of acquired knowledge as opposed to activation changes during a performance of a task.

For example, structural brain changes due to training have been observed in the cortical responses of multiple units in cats (Merzenich, 1975), rats (Kilgard and Merzenich, 1998), and adult monkeys (Recanzone et al., 1993). At a more molar level, learning-based changes in grey and white matter have been observed in human participants (see Fields, 2011; Lövdén et al., 2013; Thomas and Baker, 2013; Zatorre et al., 2012 for reviews). For example, gray and white matter changes were observed when people were trained in juggling (Draganski et al., 2004). Evidence that structural changes occur with training also comes from visuo-motor tasks, working memory tasks, and aerobics. Many of these studies involve tasks in which learning occurs as an effect of repeated training trials rather than being due to learning from instruction (as might occur in a classroom).

* Corresponding author at: Baker Hall, Department of Psychology, Center for Cognitive Brain Imaging, Carnegie Mellon University, Department of Psychology, Pittsburgh, PA 15213.
E-mail address: rmason@andrew.cmu.edu (R.A. Mason).

http://dx.doi.org/10.1016/j.neuroimage.2014.12.086
1053-8119/© 2015 Published by Elsevier Inc.

A number of tasks have shown brain-based changes in activation patterns due to training or instruction-based learning. Typically, these tasks examine brain changes in activation, in which it may be difficult to separate new representations from learned processes. Examples of tasks that examine activation in training-based learning include artificial grammars (Petersson, Folia, & Hagoort, 2012), perceptual category learning (Poldrack et al., 2001), and motor learning (Toni, Krams, Turner, & Passingham, 1998). Instruction-based learning tasks such as algebra (Anderson et al., 2012 have also resulted in activation changes. One study demonstrated both activation changes and white matter changes as a result of both instruction and repeated training in word decoding in children with dyslexia (Meyler et al., 2008; Keller & Just, 2009). Unlike these previous studies, here we look not for changes in tissues or in regional activation, but in the neural representations of specific concepts using recent methods that can identify the nature of the information that is being coded by a given fMRI activation pattern.

The neural changes in our study were assessed in terms of the multi-voxel fMRI-measured activation pattern that occurs when participants think about how a particular mechanical system works. More precisely, the study assessed how their neural representation of a system changed as they learned more about its workings. The change in knowledge about specific mechanical systems should produce measureable changes in the neural representations of those systems. Furthermore, the changes may be directly related to the content of the instruction, such that instruction that describes shared properties between systems may increase their neural similarity.

Participants were taught with a series of successive increments of information about mechanical systems. The first level of explanation provided information about the components of the mechanical systems. The second increment included limited functional information. The third increment of explanation included the entire functional and causal sequence of the components of the mechanical systems. Each of these instructional steps should result in discernable neural changes. More specifically, the progressive deepening of the explanations of the systems might be expected to produce increasing involvement of cortical systems implementing higher-level psychological processes, and unchanging involvement of lower level perceptual systems that process the visual stimulus.

Despite the absence of prior neuroimaging investigations of mechanical systems, previous research in the brain bases of general cognitive processes does provide guidance as to which cortical systems might be involved. A set of eight potential cognitive processes, which have previously been associated with cortical systems, are postulated to correspond to regions or small sets of regions (networks) involved in understanding how mechanical systems work. These eight processes (and postulated cortical systems) consist of: (1) mental animation (bilateral parietal: Boronat et al., 2005), (2) causal reasoning (right temporo-parietal and medial prefrontal: Mason and Just, 2011), (3) embodied cognition (pre- and post-central: Rueschemeyer et al., 2010), (4) semantic knowledge (left temporal: Price, 2000), (5) language in context (bilateral inferior frontal: Mestres-Missé et al., 2008), (6) biological/goal-directed motion (right temporal: Pelphrey et al., 2003), (7) rule learning (middle and superior frontal: Bunge, 2004), and (8) visual processing (occipital cortex). The contributions of these various systems might be expected to shift as the instruction and learning progresses.

The goal of this study was to examine the changes in the neural representation over the course of learning and instruction, rather than establishing the correspondence between cognitive functions and brain regions. We developed several hypotheses concerning changes in representation. First, prior to instruction, during the first exposure to only the diagram and label, the hypothesis is that the participating regions would be primarily visual in nature, loading on the occipital cortex. Subsequent neural representations should involve relatively less occipital participation. Second, following the introduction of causality information, the representation could be expected to be distributed across a large set of systems including causal inference related regions (medial frontal and right temporo-parietal) for inferring causal relations among the components' motions. Third, bilateral parietal, particularly the intraparietal sulcus, should increase in participation once components of the mechanical systems are introduced as a result of imagining components moving with respect to each other. Intuitively imagining the components moving may be a part of mental animation (Hegarty, 1992). These hypotheses provide a starting point for examining the changing involvement of cortical systems during learning.

Several recently developed methods for assessing neural knowledge representations were used in the study. One of these was the machine learning and classification of the multi-voxel activation patterns associated with each of the mechanical systems (Just et al., 2010; Mitchell et al., 2008). A related method analyzed the locations of the types of voxels whose activation levels were modulated in a consistent way by the different mechanical systems (Just et al., 2010). A third method used representational similarity analysis to assess how similarly-described systems became neurally more similar (Connolly et al., 2012). These methods can be used to converge on an account of how the neural representations change as instruction and knowledge cumulate.

Materials and methods

Participants

Fourteen college students (6 females, all right handed and native speakers of English) between the ages of 18 and 26 years ($M = 21.57$; $SD = 2.79$) participated and were included in all of the analyses (no subjects were excluded). Each participant gave signed informed consent approved by the Carnegie Mellon University Institutional Review Board. Each participant received 5 minutes of practice with the experimental paradigm on a single training item (that was not included in the experimental stimuli) before performing the task in the scanner. In a debriefing session, all participants responded positively when asked if they felt they had gained an understanding of how the mechanical systems worked. Additionally, when they were asked if they had "prior knowledge of how any of the systems worked" only one participant said he had a very basic understanding of the systems. This participant's data did not differ from the others' so it was retained in the analysis

Experimental design

In the scanner, participants were taught how four familiar devices work (a bathroom scale, a fire extinguisher, car brakes, and a trumpet). The systems were selected to vary across some potentially interesting dimensions (e.g., manipulation by hand versus foot, being composed of different types of mechanical components) as well as meeting two criteria: (1) amenability of the explanation to segmentation into successive stages; (2) informal assessment that students who were not in science or engineering would be unlikely to know how the system worked.

The experimental design consisted of the four items presented in two types of blocks: thinking (or "test") blocks and explanation (or "training") blocks. The experimental design and timing of all events (presentations, blocks and scans) are shown in Fig. 1. In the thinking blocks, which provide the main data for this study, participants were presented with each of the four items and asked to "Think about how this mechanical system might function." The explanation blocks cumulatively described how the components of each system work together to cause the system to function. During the explanation blocks, subjects were asked to read each sentence and "Think about the functioning of each stage of the mechanical system."

In the thinking blocks, each stimulus item consisted of a realistic picture of the system above a schematic diagram and a verbal label for the system, as shown in Fig. 2A. In a thinking block, the mechanical systems were presented in six presentations (i.e., repetitions) of the four systems. The presentation order of the mechanical systems in each block was randomized using a Latin square design with an additional

Please cite this article as: Mason, R.A., Just, M.A., Physics instruction induces changes in neural knowledge representation during successive stages of learning, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2014.12.086