



Published in final edited form as:

Perspect Psychol Sci. 2013 January ; 8(1): 84–87. doi:10.1177/1745691612469030.

Moving forward with fMRI data

Michael D. Rugg and

University of Texas, Dallas

Sharon L. Thompson-Schill

University of Pennsylvania

Abstract

The use of fMRI to inform cognitive theory depends upon the assumption that there is an isomorphic relationship between functional states and brain states. Even if this assumption is accepted, employing fMRI to make cognitive inferences is not straightforward. We discuss these inferential difficulties, and describe an example where fMRI data has had a significant impact on a cognitive theory despite them. According to the embodied cognition framework, accessing a conceptual feature engages the same processes that are active when the feature is directly experienced. A consistent finding from fMRI studies is that accessing a feature activates cortical regions slightly anterior to the regions activated when the feature is experienced. We suggest that this neural dissociation between accessing and perceiving a feature is inconsistent with the strong form of embodied cognition theory, and that this inconsistency would likely not have been identified without fMRI. We describe how the observation of this forward-shift has led to new ways about thinking about perceptual and conceptual representations and the relation between them. Finally, we argue that despite the strong assumptions that are the needed to make functional inferences from fMRI, this does not detract from its value as a source of convergent evidence.

On the face of it, fMRI data are ill-suited to the task of informing cognitive theory. As two neuroscientists *né* psychologists, we are both interested in drawing conclusions about cognition using fMRI data. We recognize, however, that by their nature fMRI experiments only permit causal inferences about brain states (namely, how variations in cognitive states affect the brain; Henson, 2005; Weber and Thompson-Schill, 2010). In order to reverse this inference, and draw conclusions about cognition from fMRI data, one must make strong assumptions about the relation between brain and cognition that are, as we will develop below, open to question. However, we will argue that fMRI data have nonetheless advanced at least one cognitive theory in ways that had not and arguably *could* not have occurred in their absence.

There are many reasons why one might observe dissociations and associations between mental states and brain states, not all of which inform cognitive theory. Consider the dissociation between the neural responses elicited by visual stimuli in the left versus the right visual field; the difference in brain state (which correlates with the difference in

stimulus location and the ensuing mental state) does not warrant the conclusion that, in any interesting sense, different cognitive operations are reflected in these two responses. More generally, a neural dissociation can be used to infer a functional (cognitive) dissociation only if there is an isomorphic and invariant relationship between the cognitive process of interest and the pattern of neural activity that supports the process (Rugg and Coles, 1995). Put another way, if the same cognitive process is associated with more than one pattern of neural activity then neural measures cannot easily be used to determine when the process is active and when it is not. Conversely, computationally distinct operations may be mediated by cohabiting neuronal populations (e.g., M and P cells in V1) whose differential activation is undetectable at the spatial scale afforded by fMRI. Thus, the finding that two tasks elicit co-localized (overlapping) patterns of activity does not necessarily imply that the tasks engaged functionally equivalent cognitive processes. Almost invariably, though, using fMRI to make a cognitive inference hinges upon the demonstration of either dissociable or co-localized patterns of task-related fMRI signals. But as we just noted, such inferences are valid only if there is a one-to-one relationship between the pattern of fMRI activity elicited during a task and the cognitive processes that the task engages. The two examples just given demonstrate that inferences along these lines are not necessarily valid.

The challenge raised by this special issue is to identify cases where, despite this caveat, our understanding of some aspect of cognition has been advanced with neuroimaging data. As we discuss later, although in many circumstances cognitive inferences made from fMRI data alone would be suspect, their convergence with other sources of evidence lends credence to their interpretation. However, our remit as we understand it is to identify an advance to a cognitive theory that had not (or stronger still, could not have) been made using other methods from either the psychologist's or neuroscientist's toolbox. If, in the absence of convergent evidence, one cannot rely on either dissociation or co-localization in patterns of fMRI data to make a cognitive inference, this effort would seem hopeless. We have turned to an example where the cognitive advance arises from neither a dissociation nor an association, but instead, from both.

Our argument begins in 1985, although precursors to these ideas are found in Sigmund Freud's *On Aphasia* and in embryonic form in John Locke's *An Essay Concerning Human Understanding*; but, we begin with Allport (1985), who states:

The essential idea is that the *same* neural elements that are involved in coding the sensory attributes of a (possibly unknown) object presented to eye or hand or ear also make up the elements of the auto-associated activity-patterns that represent familiar object-concepts in 'semantic memory'. This model is, of course, in radical opposition to the view, apparently held by many psychologists, that 'semantic memory' is represented in some abstract, modality-independent, 'conceptual' domain remote from the mechanisms of perception and of motor organization. (p. 53, emphasis in original).

In this seminal chapter, Allport's primary goal was to argue that information is represented as distributed patterns; however, he went a step further by proposing that, for object knowledge (i.e., "semantic memory"), these patterns in memory are shared with patterns evoked during perception. Allport sets the stage for the relevance of neural data for this

proposal by arguing that “we need also to have a model of functionally separable components *at the neural (implementation) level*” (p. 39, emphasis in original).

Today, Allport’s hypothesis is bundled into the umbrella theory of “embodied cognition”: rather than viewing cognition as a separate set of processes from those that compose our perception and action systems, we can describe cognition as the extension of these systems. Beyond mere parsimony, the embodied cognition approach to knowledge representation has the advantage of circumventing the “symbol grounding problem”, that is, the problem of how meanings become connected to the world. Embodied cognition answers this problem by connecting cognition with the world via sensory and motor processes. For this reason, the framework is appealing, and evidence for *modal* semantic memory representations is widely touted; for example, in a 2003 review, Barsalou and colleagues summarize the state of the field with, “these empirical results and theoretical analyses implicate modality-specific systems in the representation and use of conceptual knowledge” (p. 84). What are the empirical results? There are several reviews of relevant findings (e.g., Barsalou et al. 2003; Martin, 2007; Thompson-Schill, 2003), and we offer here one brief summary of the neuroimaging data that have been argued to support embodied cognition accounts:

Conceptualizing the colors of objects activates a brain area near V4, the area that processes color sensations. Conceptualizing object form activates areas near brain centers for perceiving form. Conceptualizing object motion activates areas near brain centers for perceiving motion. (Barsalou et al., 2003, p. 87).

This quotation, which comes from a review championing the embodied cognition account, contains an important word, in each of three consecutive sentences, that is the basis for the argument we develop below: “*near*”. ‘Near’ is not ‘same’. Thus, one could argue that these fMRI findings, often marshaled to defend embodied theories of concepts, could be said to refute them.

Returning to our charge, have fMRI data contributed to our understanding of concept representation? As we argued above, dissociable patterns of activation cannot necessarily be used to infer dissociable cognitive processes, because the same process might be subserved by multiple neural populations. Put in these terms, the statement in Allport (1985), found in the excerpt above, makes the erroneous assumption that there is an isomorphism between “neural elements” and cognitive processing. But without that assumption, we should conclude nothing about the status of embodied cognition theories from these data. Similarly, the co-localization of activation would tell us nothing about the grounding of meaning in sensorimotor systems since we cannot be confident that the co-localization reflects activation of the same neural populations (cf. Mahon & Caramazza, 2008). Yet, we suggest that the observation of *near-but-not-same* activation patterns in memory and perception (or action) has indeed moved the field forward.

We make our argument in the domain of color, but it could in principle apply to form, motion, or action. Much of the seminal work in the retrieval of color knowledge was conducted by Alex Martin and his colleagues, who were the first to compare the fMRI response to color perception and color knowledge retrieval in the same subjects (Chao & Martin, 1999): whereas passive viewing of chromatic stimuli activated the lingual gyrus,

naming the colors of common objects activated a more anterior region in the fusiform gyrus. Near, but not the same. In response to the challenge that this finding creates for embodied cognition theories, Simmons and colleagues (2007) compared fMRI responses during color perception and color memory retrieval using a more demanding perceptual task (Beauchamp et al., 1999). They observed color-selective activity during perception in both posterior lingual *and* anterior fusiform regions, the latter overlapping with the activity evoked during color memory retrieval. In other words, by changing the processing demands of the perception task, they changed the resulting pattern of activation, such that it partially overlapped with the ‘anteriorly-shifted’ activity elicited during retrieval of color information from memory.

Simmons et al. (2007) discussed these two regions with regard to levels of color representation, positing that there are low- and high-level representations, supported by different anatomical substrates (distributed along the posterior-anterior axis), that contribute differentially to perception and memory. Hsu et al., (2011) adopted a complementary strategy, varying the memory task rather than the perception task. These authors manipulated the color information required to perform the task (high fidelity color representations to determine that a strawberry is a closer shade of red to a tomato than it is to a raspberry versus lower fidelity color representations sufficient to determine that a strawberry is a closer color match to a cherry than to a lemon). Hsu et al. also measured subjects’ self-reported tendency to prefer visualizing rather than verbalizing information. They found that both task and subject factors affected the amount of overlap between the activity elicited during the memory and perception tasks.

The finding that the overlap between cortical regions involved in color perception and color memory is task-dependent, is, to our knowledge, unique to fMRI studies. Indeed, with the possible exception of the closely related but largely superseded method of positron emission tomography (PET), it is arguable that no other method possesses the combination of spatial resolution and coverage necessary to identify this phenomenon. In particular, it is difficult to imagine analogous findings emerging from a lesion-deficit study, given the very high spatial specificity that would be required of the lesions.

We argue that, together, the findings of Simmons et al. (2007) and Hsu et al. (2011) have contributed in an important way both to our knowledge of the relation between color memory and color perception and to embodied cognition theories more generally, reconciling the predictions of Allport (1985) with the “anterior shift” from perception to memory observed in early fMRI reports. The findings lead to the hypothesis that a given feature or property (such as color) is represented at multiple levels of abstraction (such that abstraction increases in more anterior regions), and both perception and memory can tap into these varying representations. Although there might be a tendency for perceptual tasks to recruit less abstract representations and for memory tasks to recruit more abstract representations, both task and subject factors can influence which representations are involved in memory and perception (and might also influence whether a phenomenological experience of “imagery” accompanies memory retrieval).

Critically, this advance hinged neither on an association between activation patterns (i.e., co-activation during perception and memory) nor on a dissociation (differential activation between the two), but rather on the observation that the extent of association or dissociation can be systematically manipulated. This observation is crucial. That a dissociation can be ‘transformed’ into an association by varying the specificity of the information required to perform a task rules out several theoretically uninteresting interpretations of the anterior shift. In particular, this finding is inconsistent with an interpretation of the shift in terms of differences in neural activity that merely reflect whether input to a functionally homogeneous cortical region is ‘bottom-up’ or ‘top-down’. When Simmons et al. (2007) reported the first evidence of direct overlap between memory and perception, they suggested that “it will be important for future research to articulate the commonalities and boundaries between perception and knowledge, if we are to develop satisfactory process-level accounts of these fundamental cognitive abilities” (p. 2809). We argue that fMRI data have done just that

Although we have identified a case where it can be argued that fMRI has uniquely advanced a cognitive theory, we acknowledge that such cases are rare. We therefore conclude by asking how fMRI can be employed to inform cognitive theory more generally. We have already noted that cognitive inferences based on patterns of dissociations or associations in fMRI data rest on strong and (at present) unverifiable assumptions. This is no less true however of any other methodology, including those – such as the study of healthy or brain-lesioned individuals - that rely on behavioral rather than neural measures. We argue that the value of fMRI as a tool to inform cognitive theory is best realized when findings from the method are integrated with findings from other methods that, while no less assumption-bound, depend on *different* assumptions. In short, we are making the mundane point that convergent evidence from multiple methods is preferable to evidence obtained with any single method. From this perspective, the assumptions underlying the use of fMRI to make inferences about cognitive theory are much less daunting than when the method is viewed in isolation.

Acknowledgements

Research described in this paper was supported by National Institutes of Health grants R01MH70850 and R01EY021717 awarded to STS. We thank Nina Hsu for comments on a draft version of the paper.

References

- Allport, DA. Distributed memory, modular subsystems and dysphasia. In: Newman, SK.; Epstein, R., editors. *Current perspectives in dysphasia*. New York: Churchill Livingstone; 1985. p. 32-60.
- Barsalou LW, Simmons WK, Barbey AK, Wilson CD. Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*. 2003; 7:84–91. [PubMed: 12584027]
- Beauchamp MS, Haxby JV, Jennings JE, DeYoe EA. An fMRI version of the Farnsworth-Munsell 100-Hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cerebral Cortex*. 1999; 9:257–263. [PubMed: 10355906]
- Henson R. What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology*. 2005; 58:193–233. [PubMed: 15903115]
- Hsu NS, Kraemer DJM, Oliver RT, Schlichting ML, Thompson-Schill SL. Color, context, and cognitive style: Variations in color knowledge retrieval as a function of task and subject variables. *Journal of Cognitive Neuroscience*. 2011; 23:2544–2557. [PubMed: 21265605]

- Martin A. The representation of object concepts in the brain. *Annual Review of Psychology*. 2007; 58:25–45.
- Mahon BZ, Caramazza A. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology (Paris)*. 2008; 102:59–70.
- Rugg, MD.; Coles, MGH. The ERP and cognitive psychology: conceptual issues. In: Rugg, MD.; Coles, MGH., editors. *Electrophysiology of Mind*. New York: Oxford University Press; 1995. p. 27-39.
- Simmons WK, Ramjee V, Beauchamp MS, McRae K, Martin A, Barsalou LW. A common neural substrate for perceiving and knowing about color. *Neuropsychologia*. 2007; 45:2802–2810. [PubMed: 17575989]
- Thompson-Schill SL. Neuroimaging studies of semantic memory: Inferring “how” from “where”. *Neuropsychologia*. 2003; 41:280–292. [PubMed: 12457754]
- Weber MJ, Thompson-Schill SL. Functional neuroimaging can support causal claims about brain function. *Journal of Cognitive Neuroscience*. 2010; 22:2415–2416. [PubMed: 20201629]