How fMRI can inform cognitive theories

Mara Mather,
University of Southern California

John T. Cacioppo, and
University of Chicago

Nancy Kanwisher
Massachusetts Institute of Technology

The goal of cognitive psychology is to develop and test theories about how the mind works. In this commentary, we address the question of how fMRI can be used to help cognitive psychologists understand cognition. We start by putting forth our own views: that fMRI can inform theories about cognition by helping to answer at least four distinct types of questions.

Question 1: Which (if any) functions can be localized to specific brain regions?

Coltheart (this issue) argues that, "we do not have any evidence showing that any particular form of brain activation is a sign that some particular type of cognitive operation is being performed" (p. XX). The degree to which the brain is composed of modules that each carry out a specific aspect of cognition or of components that each participate in many different processes has been hotly debated for years (for review see Kanwisher, 2010). However, evidence indicates that some cortical regions respond selectively to certain categories of visual stimuli (for instance, the fusiform face area, the parahippocampal place area and the extrastriate body area; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006). This category selectivity bears directly on one of the basic tasks of cognitive psychology, which is to come up with an inventory of dissociable mental processes. Thus, the finding that a given cortical region is selectively engaged in a particular mental process can be informative not because it tells us the *location* of that process (why would a psychologist care?), but because it suggests that the brain, and hence the mind, contains specialized mechanisms for that particular mental process. As reviewed below, localization findings have also been used in many creative ways to test other hypotheses about cognition.

Question 2: Can markers of mental process X be found during task Y?

From the perspective of some cognitive psychologists, the focus of initial fMRI research on localization did not appear to provide information about how the mind worked (for a review see Shimamura, 2010). For instance, Uttal (2001) argues that, "Even if we could associate precisely defined cognitive functions in particular areas of the brain (and this seems highly unlikely), it would tell us very little if anything about how the brain computes, represents, encodes, or instantiates psychological processes (Uttal, 2001, p. 217).

But to say that neuroimaging answers only "where" questions is to confuse the superficial format of raw neuroimaging data with the content of the questions those data can answer; Neuroimagers collecting fMRI data need no more restrict themselves to "where" questions

than cognitive psychologists measuring reaction times need limit themselves to "when" questions.

Indeed, looking back over the past two decades of work using fMRI, one can see that an initial focus on "brain mapping" provided a critical foundation for the field that allowed it to go beyond "where" questions to address "how" questions. To the extent that highly specialized brain regions are discovered, and their functional specialization well established[1], activity in these regions can serve as markers for specific cognitive functions, enabling us to ask whether and to what extent mental process X is engaged in task Y. For example, by showing people a series of scenes and faces and asking them to remember one type of stimuli and ignore the other during a retention interval, researchers tested the hypothesis that people suppress mental representations of distracting to-be-ignored stimuli (Gazzaley, Cooney, McEvoy, Knight, & D'Esposito, 2005) and that older adults show less suppression of the distracting stimuli (Gazzaley, Cooney, Rissman, & D'Esposito, 2005). In this study, activity in the fusiform face area (for faces) or parahippocampal place area (for scenes) was used as a signal of enhancement or suppression of processing that type of stimuli.

Another example comes from work testing the hypothesis that objects are the units of attentional selection, rather than locations or features (O'Craven, Downing, & Kanwisher, 1999). Here, each stimulus in a sequence consisted of a face transparently superimposed on a house, with either the face or the house moving. Participants had to monitor either repetitions in faces or houses or the direction of motion. Attending to one attribute of an object (e.g., the motion of a house) enhanced activity not only in the region specialized for that attribute (e.g., area MT/MST for motion) but also in the region specialized for its other attribute (e.g., the parahippocampal place area for houses) compared against the other object's attributes. These results provide evidence for object-based attention and argue against locations or features being the unit of selection.

Question 3: How distinct are the representations of different stimuli or tasks?

One of the central tasks of cognitive psychology is to characterize mental representations. The crux of the characterization of a mental representation is the specification of its invariant and equivalence classes: which entities does it take to be the same, and which different? A representation common to two different viewpoints of an object is an invariant representation of object shape, whereas a representation that is common to the word "dog" and a picture of a dog is an abstract semantic representation. This is the fundamental logic behind priming methods in cognitive psychology and looking time methods in cognitive development research. fMRI has two methods that follow the same basic logic, enabling us to characterize neural representations in particular brain regions by asking which stimuli are treated as different and which the same: fMRI adaptation (Grill-Spector & Malach, 2001) and fMRI multivariate pattern analysis (Haxby, 2012; Norman, Polyn, Detre, & Haxby, 2006).

Multivariate pattern analysis is sensitive enough to decode differences in the visual stimuli being viewed, such as the orientation of a striped pattern (Haynes & Rees, 2005; Kamitani & Tong, 2005), the movement direction in a field of dots (Kamitani & Tong, 2006), the semantic category of a word (Mitchell et al., 2004), the category of object (e.g., animals,

---

[1]Poldrack has correctly criticized "reverse inference", in which mental processes are inferred from the location of fMRI responses (Poldrack, 2006). Note however that for the few cases of brain regions whose functions are very specific, and for which that specificity is widely replicated, these regions can indeed support solid reverse inferences.

cars, planes), and in some cases even the exemplar of a category (e.g., cows, frogs, turtles) (Cichy, Chen, & Haynes, 2011). For instance, representations of scenes in the PPA (Epstein & Kanwisher, 1998) have been shown to encode the layout of space in a scene (Kravitz, Peng, & Baker, 2011; Park, 2011 #7414, and to generalize across photographs and line drawings of scenes, Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011), whereas representations in early visual cortex encode distance, and neither encode conceptual/ semantic information about scenes (Kravitz et al., 2011). In contrast, lateral occipital cortex encodes information about the content of scenes, such as whether it is an urban or natural setting (Park, Brady, Greene, & Oliva, 2011).

Probing these types of representations can also address theoretical questions about the structure of cognitive processes. For instance, one study tested the hypothesis that information in working memory is maintained by the same sensory regions that process those stimuli when they are first perceived (Serences, Ester, Vogel, & Awh, 2009). This study showed participants colored grating patterns and asked them to either remember the color or the grating orientation over a 10-second delay. When participants were trying to remember the color, multivariate pattern analyses were able to classify the color but not the orientation of the stimulus being maintained based on activity in primary visual cortex during the delay period. In contrast, when participants were trying to remember the orientation, orientation but not color could be classified. These results suggest that the sustained stimulus-specific patterns in primary visual cortex reflect active maintenance in working memory. These and other fMRI data challenge models that assume that working memory requires dedicated buffers or storage sites (for a review see D'Esposito, 2007).

Question 4: Do two tasks X and Y engage common or distinct processing mechanisms?

A fourth way that fMRI can inform cognition does not require any commitment to or prior finding of strong functional specificity of a particular region of the brain. Specifically, fMRI provides a natural way to ask one of the classic questions of cognitive psychology: do two tasks X and Y engage common or distinct processing mechanisms? If conducted properly, experiments showing overlapping brain activation for the two tasks, with appropriate control conditions, within individual subjects, can provide evidence for common mechanisms. For example one study found common brain regions engaged across diverse forms of visual attention (Wojciulik & Kanwisher, 1999), arguing that these diverse attentional phenomena have more in common than their name. Conversely, other studies have shown that distinct mechanisms are engaged in high-level language processing but not in deductive logic (Monti, Parsons, & Osherson, 2009), algebra (Monti, Parsons, & Osherson, 2012), or other cognitive phenomena including music (Fedorenko, Behr, & Kanwisher, 2011; Fedorenko, McDermott, Norman-Haignere, & Kanwisher, in press) – demonstrating powerful dissociations between language and (many aspects of) thought. Another example comes from studies testing the hypothesis that there is a unified attentional bottleneck involved in both perception and decision-making (Jiang & Kanwisher, 2003; Tombu et al., 2011). These studies found that the same brain regions showed "bottleneck" properties during a perceptual encoding task and during a speeded response decision task; these findings are inconsistent with task switching accounts that posit independent substrates for encoding and response selection.

Another question that has been tackled by examining the similarity of brain activation patterns across two conditions is whether honesty results from the absence of temptation or from the active resistance of temptation. Individuals given the opportunity to cheat who were honest showed no greater activity in brain regions associated with behavior control than those who were in a control condition with no opportunity to cheat (or any differences

in activity in any other brain regions, even at a low threshold), whereas those who were dishonest showed significantly more activity in brain regions associated with cognitive control (Greene & Paxton, 2009). The similarity of brain activation patterns among those who were honest and those with no opportunity to cheat argues against the hypothesis that honesty results from the active resistance of temptation—at least in the context of the task used in that study.

The observation that two tasks that have been thought of as distinct engage common processing mechanisms can also generate new theoretical predictions. For instance, fMRI research that directly compared imagining the future to remembering the past has revealed that these processes engages the same neural networks (Buckner & Carroll, 2007; Schacter & Addis, 2007; Szpunar, Watson, & McDermott, 2007). This surprising finding led to the novel hypothesis that older adults, who have impaired function in the neural networks involved in remembering past events, should also show impairments in mentally simulating the future. Subsequent research revealed that, indeed, older adults generate fewer distinct details when simulating the future and that this deficit is correlated with their relational memory abilities and with how many distinct details they generate when remembering the past (Addis, Wong, & Schacter, 2008).

How effectively has fMRI contributed to cognitive theory?

So far, we have outlined four ways in which fMRI provides tools that *can* be used to address theoretical debates about cognition. But how effectively have these tools been used? As mentioned by Coltheart in this issue, a survey of fMRI studies investigating cognitive functions published between 2007 and 2011 in eight journals found that most (89%) of the studies aimed to localize cognitive processes while 11% aimed to test a theory of cognition (Tressoldi, Sella, Coltheart, & Umilta, 2012). Of those that aimed to test a cognitive theory, Tressoldi et al. argued that around half committed the "consistency fallacy" (Mole & Klein, 2010), as these studies apparently claimed that their study supports some theory because the data from that study are consistent with the theory without also providing evidence that the study could have yielded a pattern of data that would have opposed that theory. Based on these metrics, only a small set of fMRI studies of cognition test cognitive theories.

Does this low hit rate mean that fMRI has properties that make it less likely to be able to inform cognitive theories than behavioral data? For instance, Uttal (2001, p. 206) argues that behaviorism approaches are better suited for the study of psychological processes than neuroimaging approaches; "One other general goal of this book has been to champion the resurrection of an underappreciated, yet scientifically sounder, approach to the study of psychological processes—behaviorism." So how does the "cognitive theory" track record of fMRI stack up against that of behavioral methods? Although, to our knowledge, no one has attempted this type of comparison, the case can be made that resolving controversies between theories in cognitive psychology is surprisingly difficult regardless of the methods used (Greenwald, 2012). Of 13 examples of prominent controversies in cognitive and social psychology originating in the 1950's through the early 1980's reviewed by Greenwald, only one—whether mental rotation required visual representation or not—was clearly resolved. Interestingly, it was not resolved by behavioral methods but by neuroimaging (Ganis, Keenan, Kosslyn, & Pascual-Leone, 2000; Kosslyn, Digirolamo, Thompson, & Alpert, 1998).

What is needed for fMRI research to inform cognitive theory?

In this issue, Wixted and Mickes state that "fMRI can inform cognitive theories that make predictions about patterns of activity in the brain" (p. XX). The implication, as made explicit by Coltheart, is that cognitive theories that do not make predictions about the results of

functional neuroimaging experiments cannot be informed by the results of neuroimaging experiments. We argue that, to the contrary, results from fMRI studies can be used to help resolve theoretical debates about cognition even when the theories involved make no predictions about the brain regions involved. An example described above is the fMRI study testing the hypothesis that objects are the units of attentional selection, rather than locations or features (O'Craven et al., 1999). In that study, the fMRI results were able to distinguish between processing of two different objects superimposed in the same location and their features (in this case, motion). The specific brain regions involved in processing each type of object and the motion were not relevant for testing the theory – but knowledge about these regions allowed the researchers to decode which type of information was being processed. As in this example, researchers have been using the decoding aspects of fMRI in many creative ways to test research questions that themselves have little direct connection to the brain. For instance, do older adults show more impairment in reflective selective attention than in perceptual selective attention (Mitchell, Johnson, Higgins, & Johnson, 2010)? Does a deficit in processing multiple stimuli at once contribute to older adults' binding deficits (Chee et al., 2006)?

Coltheart argues that cognitive systems need to be defined first before neuroimaging data are collected, hence that neuroimaging is only informed by, rather than informing, cognition. However, rarely does or should any area of science lock in its definitions, then just collect data while insisting that those definitions remain rigidly fixed in place. Rather, interesting experimental work alters our understanding of the very phenomena we set out to explore. Coltheart's argument says not that cognition cannot be informed by the brain, but rather that it should not allow itself to be. But why not let our ideas about cognition be shaped in part by neuroimaging data? For instance, as described above, having discovered that imagining the future activates the same neural networks as remembering the past, researchers who knew that those neural networks were impaired in older adults were able to generate a new prediction about behavior—and indeed demonstrated that older adults imagine the future less vividly than younger adults do (Addis et al., 2008).

In sum, we disagree with Coltheart that imaging can only inform cognitive theories when we can assume that "cognitive process C is implemented in brain region X and nowhere else in the brain, and brain region X subserves cognitive process C and no other process" (p. XX). Instead, as outlined above, there are many ways that imaging can inform cognition without meeting this high bar.

The importance of having cognitive theories inform neuroimaging research

We began by acknowledging the goal of cognitive psychology as developing and testing theories about how the mind works. This is not the only worthwhile goal, however. With the advent of neuroimaging techniques, cognitive psychologists (and psychological scientists more generally) have also begun to advance the goal of developing and testing theories about how the *brain* works (Cacioppo & Decety, 2009). As evidence, Wixted and Mickes (this issue) note that many fMRI studies rely on a cognitive theory to interpret their results, even when they are not testing the theory. Psychological theory is critically important to understanding the human brain and also in designing experiments that differentiate different cognitive processes. Given the complexity of the human brain, progress in understanding its functional organization and structure depends on sophisticated theoretical specifications of the psychological representations and processes that differentiate two or more comparison conditions. Psychological scientists, therefore, are well positioned to lead the search for brain mechanisms underlying psychological processes. Doing so constitutes an expansion of the purview of psychological science beyond a science of behavior, beyond a science of the mind, to include a science of the brain (Cacioppo & Decety, 2009).

Concluding comments

In the 20 years since the first publication using fMRI, there has been an explosion of interest in this method and many have employed it to investigate questions about cognition. In this commentary, we outlined four ways in which we believe that fMRI results can inform our understanding of cognition. First, it can answer questions about which functions can be localized to specific brain regions, questions that are of critical interest for those examining issues related to the modularity of the brain (e.g., Blumstein, Cabeza & Moscovitch, Chiao & Immordino-Yang). Second, fMRI data can be used as markers of particular mental processes, allowing insight into what processes are being engaged during different tasks. Third, fMRI can answer questions about exactly what information is represented in each region of the brain. Such data can be used to address theoretical questions about the nature of memory reactivation (e.g., Levy & Wagner) and working memory (e.g., Reuter-Lorenz) as well as basic questions about the structure of cognitive processes (e.g., Serences et al., 2009). Fourth, fMRI can answer questions about whether two tasks engage common or distinct processing mechanisms. This strategy can provide important evidence to address theoretical questions about the nature of tasks (e.g., Rugg & Thompson-Schill) and how functional circuitry reorganizes with age (e.g., Park & McDonough).

Of course, powerful as it is, fMRI is impotent to answer some questions. First, fMRI can never address the causal role of a particular brain region in a particular task, though showing correlations between fMRI signals and behavior help somewhat. Definitive answers to questions of causal role thus require other methods such as transcranial magnetic stimulation, electrical brain stimulation, and studies of patients with brain damage. Second, fMRI for the most port does not have the necessary temporal resolution to reveal the workings of thought, the component stages of which generally proceed on the scale of tens or hundreds of milliseconds, not seconds. Precise timing information from the human brain is available from ERP and MEG, intracranial recordings, and TMS. Third, even with high-resolution fMRI and even with sophisticated pattern analysis methods, each voxel pools neural activity over hundreds of thousands of neurons, so the signal we can see with fMRI is a drastically subsampled version of the actual language in which neurons talk to each other. This means that we cannot tell whether the same neurons are involved even when activity looks the same across multiple conditions. Finally, like other neuroscience methods, even the most astonishing demonstration of a specific neural representation in a given region of the brain leaves open the question of whether that representation was computed locally in the region where it is observed, or whether instead that information was inherited from an earlier stage of processing. The best approach to answering questions about cognition therefore is a synergistic combination of behavioral and neuroimaging methods, richly complemented by the wide array of other methods in cognitive neuroscience.

References

Addis DR, Wong AT, Schacter DL. Age-related changes in the episodic simulation of future events. Psychological Science. 2008; 19:33–41. [PubMed: 18181789]

Buckner RL, Carroll DC. Self-projection and the brain. Trends in Cognitive Sciences. 2007; 11:49–57. [PubMed: 17188554]

Cacioppo JT, Decety J. What Are the Brain Mechanisms on Which Psychological Processes Are Based? Perspectives on Psychological Science. 2009; 4:10–18.

Chee MWL, Goh JOS, Venkatraman V, Tan JC, Gutchess A, Sutton B, et al. Age-related changes in object processing and contextual binding revealed using fMR adaptation. Journal of Cognitive Neuroscience. 2006; 18:495–507. [PubMed: 16768356]

Cichy RM, Chen Y, Haynes JD. Encoding the identity and location of objects in human LOC. Neuroimage. 2011; 54:2297–2307. [PubMed: 20869451]

D'Esposito M. From cognitive to neural models of working memory. Philosophical Transactions of the Royal Society B: Biological Sciences. 2007; 362:761–772.

Downing PE, Chan AWY, Peelen MV, Dodds CM, Kanwisher N. Domain specificity in visual cortex. Cerebral Cortex. 2006; 16:1453–1461. [PubMed: 16339084]

Epstein R, Kanwisher N. A cortical representation of the local visual environment. Nature. 1998; 392:598–601. [PubMed: 9560155]

Fedorenko E, Behr MK, Kanwisher N. Functional specificity for high-level linguistic processing in the human brain. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:16428–16433. [PubMed: 21885736]

Fedorenko E, McDermott JH, Norman-Haignere S, Kanwisher N. Sensitivity to musical structure in the human brain. Journal of Neurophysiology. (in press).

Ganis G, Keenan JP, Kosslyn SM, Pascual-Leone A. Transcranial magnetic stimulation of primary motor cortex affects mental rotation. Cerebral Cortex. 2000; 10:175–180. [PubMed: 10667985]

Gazzaley A, Cooney JW, McEvoy K, Knight RT, D'Esposito M. Top-down enhancement and suppression of the magnitude and speed of neural activity. Journal of Cognitive Neuroscience. 2005; 17:507–517. [PubMed: 15814009]

Gazzaley A, Cooney JW, Rissman J, D'Esposito M. Top-down suppression deficit underlies working memory impairment in normal aging. Nature Neuroscience. 2005; 8:1298–1300.

Greene JD, Paxton JM. Patterns of neural activity associated with honest and dishonest moral decisions. Proceedings of the National Academy of Sciences. 2009; 106:12506–12511.

Greenwald AG. There Is Nothing So Theoretical as a Good Method. Perspectives on Psychological Science. 2012; 7:99–108.

Grill-Spector K, Malach R. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychologica. 2001; 107:293–321. [PubMed: 11388140]

Haxby JV. Multivariate pattern analysis of fMRI: The early beginnings. Neuroimage. 2012; 62:852–855. [PubMed: 22425670]

Haynes JD, Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nature Neuroscience. 2005; 8:686–691.

Jiang YH, Kanwisher N. Common neural mechanisms for response selection and perceptual processing. Journal of Cognitive Neuroscience. 2003; 15:1095–1110. [PubMed: 14709229]

Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. Nature Neuroscience. 2005; 8:679–685.

Kamitani Y, Tong F. Decoding Seen and Attended Motion Directions from Activity in the Human Visual Cortex. Current Biology. 2006; 16:1096–1102. [PubMed: 16753563]

Kanwisher N. Functional specificity in the human brain: A window into the functional architecture of the mind. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:11163–11170. [PubMed: 20484679]

Kosslyn SM, Digirolamo GJ, Thompson WL, Alpert NM. Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography. Psychophysiology. 1998; 35:151–161. [PubMed: 9529941]

Kravitz DJ, Peng CS, Baker CI. Real-world scene representations in high-level visual cortex: It's the spaces more than the places. Journal of Neuroscience. 2011; 31:7322–7333. [PubMed: 21593316]

Mitchell KJ, Johnson MR, Higgins JA, Johnson MK. Age differences in brain activity during perceptual versus reflective attention. Neuroreport. 2010; 21:293–297. [PubMed: 20125054]

Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X, Just M, et al. Learning to Decode Cognitive States from Brain Images. Machine Learning. 2004; 57:145–175.

Mole C, Klein C. Confirmation, refutation and the evidence in fMRI. Foundational issues in human brain mapping. 2010:99–112.

Monti MM, Parsons LM, Osherson DN. The boundaries of language and thought in deductive inference. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:12554–12559. [PubMed: 19617569]

Monti MM, Parsons LM, Osherson DN. Thought beyond language: Neural dissociation of algebra and natural language. Psychological Science. 2012; 23:914–922. [PubMed: 22760883]

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in Cognitive Sciences. 2006; 10:424–430. [PubMed: 16899397]

O'Craven KM, Downing PE, Kanwisher N. fMRI evidence for objects as the units of attentional selection. Nature. 1999; 401:584–587. [PubMed: 10524624]

Park S, Brady TF, Greene MR, Oliva A. Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. Journal of Neuroscience. 2011; 31:1333–1340. [PubMed: 21273418]

Poldrack RA. Can cognitive processes be inferred from neuroimaging data? Trends in Cognitive Sciences. 2006; 10:59–63. [PubMed: 16406760]

Schacter DL, Addis DR. The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philosophical Transactions of the Royal Society B-Biological Sciences. 2007; 362:773–786.

Serences JT, Ester EF, Vogel EK, Awh E. Stimulus-Specific Delay Activity in Human Primary Visual Cortex. Psychological Science. 2009; 20:207–214. [PubMed: 19170936]

Shimamura AP. Bridging Psychological and Biological Science: The Good, Bad, and Ugly. Perspectives on Psychological Science. 2010; 5:772–775.

Szpunar KK, Watson JM, McDermott KB. Neural substrates of envisioning the future. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:642–647. [PubMed: 17202254]

Tombu MN, Asplund CL, Dux PE, Godwin D, Martin JW, Marois R. A unified attentional bottleneck in the human brain. Proceedings of the National Academy of Sciences. 2011; 108:13426–13431.

Tressoldi PE, Sella F, Coltheart M, Umilta C. Using functional neuroimaging to test theories of cognition: A selective survey of studies from 2007 to 2011 as a contribution to the Decade of the Mind Initiative. Cortex; a journal devoted to the study of the nervous system and behavior. 2012; 48:1247–1250.

Uttal, WR. The new phrenology: The limits of localizing cognitive processes in the brain. Cambridge, MA: MIT Press; 2001.

Walther DB, Chai B, Caddigan E, Beck DM, Fei-Fei L. Simple line drawings suffice for functional MRI decoding of natural scene categories. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:9661–9666. [PubMed: 21593417]

Wojciulik E, Kanwisher N. The generality of parietal involvement in visual attention. Neuron. 1999; 23:747–764. [PubMed: 10482241]