



doi:10.1093/scan/nsl019

SCAN (2006) 1 of 3

Studying mind and brain with fMRI

Marc G. Berman, John Jonides, and Derek Evan Nee

University of Michigan, USA

The explosion in publications using functional magnetic resonance imaging (fMRI) warrants an examination of how the technique is being used to study processes of mind and brain. Here, we propose a classification of fMRI studies that reveals how this technique is being used in the service of understanding psychological and neural processes and the relationship between the two.

In 1993, the number of published articles citing functional magnetic resonance imaging (fMRI) was fewer than 20. In 2003, that number was nearly 1800. Although not exponential, the rise in publications year by year during this interval was positively accelerated. What is it that psychologists and neuroscientists have found useful about this technique (and about other techniques such as positron emission tomography and event-related potentials)? That is, what use is being made of data from fMRI experiments? As this technology matures, we take a moment to reflect on the ways in which fMRI has been used to study mind and brain.

We implement this reflection by attempting to classify the approaches that authors have used in applying fMRI data to understanding psychological and neural mechanisms. By our reckoning, the number of approaches taken by authors is quite small. As a representative sample, we surveyed all primary empirical studies using fMRI, published in the journals *Science* and *Nature* from 2000–2006. There are 64 such studies, and our analysis of them indicates that they adopted one or more of the approaches that we describe in subsequent sections.

STUDIES OF LOCALIZATION

One approach to the use of fMRI is motivated by an interest in localizing psychological functions to brain regions. The intent of authors who adopt this approach is to identify brain behavior correlations—that is, to discover how psychological processes are localized in brain tissue. There is substantial value in understanding localization, both to understand the normal organization of modules of processing and to predict the nature of deficits that will arise when brain tissue is damaged.

One example of this approach comes from Downing *et al.* (2001) who were motivated to discover whether object recognition makes use of the same neural machinery

regardless of the object being recognized, or whether there are modules of processing tailored to specific classes of objects. Specifically, these authors were concerned with mapping out the brain regions responsible for recognition of parts of the human body, a finding that might help inform us about deficits such as those documented by Shelton *et al.* (1998), having to do with failures to process semantic information about body parts. They discovered remarkable consistency among participants in the activation of a region in right lateral occipitotemporal cortex that responded more strongly to images of human bodies than to other classes of objects, consistent both with the neuropsychological evidence in humans and with the records of single-cell activity in monkeys for similar material. Thus, a program of this sort is of value not only in mapping out the architecture of the visual processing stream, but also in helping to understand neuropsychological pathologies that involve selective deficits in processing certain classes of objects.

One issue that arises in the study of localization is just how modular brain organization is. That is, is there a one-to-one mapping of functions onto brain regions? Cases such as face recognition (e.g. Kanwisher, 2000) and recognition of parts of the human body (Downing *et al.*, 2001) suggest that there may be such a straightforward mapping, but the work of Haxby *et al.* (2001) reveals that the coding is more complex than this. They proposed a model of object recognition in which the processing of faces and other objects is distributed over a swath of brain regions. According to their model, it is the *pattern* of activation over the regions critical to object identification that is critical to object recognition, and not individual, encapsulated brain areas that are activated selectively for different stimuli. The authors found uniquely distributed patterns of neural activity in ventral temporal cortex (VTC) for the identification of faces, houses, cats and various man-made objects. Activation patterns in VTC predicted the category of object being viewed with 96% accuracy, showing that objects are uniquely represented in VTC. This result is consistent with the regional specificity implied by the work of Downing *et al.* (2001). What is striking is that even when the region that responded maximally to a class of objects (e.g. faces) was

Preparation of this manuscript was supported in part by grant BCS 0520992 from the National Science Foundation (NSF) to the University of Michigan and by NSF Graduate Fellowships to M.G.B. and D.E.N.

Correspondence should be addressed to Marc G. Berman, Department of Psychology, University of Michigan, Ann Arbor, MI 48109-1109, USA.

**NOT FOR
PUBLIC RELEASE**

removed from the analysis, the pattern of activation in the remaining regions in VTC still correctly discriminated the classes of objects with 94% accuracy. That is, while VTC may be the site of processes critical to visual object recognition, its organization is not strictly modular, with an overlapping and distributed representation of objects appearing to be the most apt characterization. These results show that a program of research concerned with localization need not be restricted to identifying one-to-one brain-to-behavior mappings.

STUDIES OF COMMONALITIES IN BRAIN ACTIVATION

A corollary to studies of localization is this: If two tasks lead to activation of common brain areas, then these two tasks or behaviors are likely to share some process or processes (Jonides *et al.*, 2006; Henson, 2006). As Poldrack (2006) and Coltheart (2006) demonstrated, this logic is not infallible. Even so, examining cases in which regional brain activation from one task is coextensive with activation from another task can be quite informative. For example, Eisenberger *et al.* (2003) showed that neural activity in anterior cingulate and right ventral prefrontal cortex found during the experience of social exclusion (social pain) was very similar to that found during the experience of physical pain, suggesting that the experiences corresponding to these two types of pain are quite similar. Another example comes from the work of Wager *et al.* (2004) who studied the neural mechanisms underlying the placebo effect. Wager *et al.* (2004) found that placebo analgesia was related to decreased activity in regions sensitive to physical pain such as the thalamus, insula and anterior cingulate cortex (Wager *et al.*, 2004). That is, there was an overlap in the regions that were decreased in activation by placebos with the regions that are increased in activation by physical pain. This result leads to the hypothesis that placebos exercise their effect by lowering the activation in brain regions that respond to physical pain, thereby exercising their analgesic effect on central processing mechanisms. These studies demonstrate that fMRI can be used to infer the cognitive processes involved in one task by showing similarities in brain activation to a better understood task.

STUDIES OF DISTINCTIVENESS IN BRAIN ACTIVATION

The complement to studies of common brain activations are studies that seek to discover distinctive activations between two tasks. Discovering such dissociations permits the inference that two tasks have different cognitive processes mediating them (e.g. Smith and Jonides, 1995; Smith *et al.*, 1996; Jonides *et al.*, 2006; Henson, 2006; Poldrack, 2006). Thus, studies of distinctive activations when added to studies of common activations enable a program of research that will gradually build an architecture of psychological processing out of an architecture of brain activity. Of course, one caveat with this technique is that

most findings of distinctive activations yield results of partial overlap in activations, so the distinctiveness that is found may be quantitative rather than a qualitative one.

An excellent example of a search for dissociations comes from the work of MacDonald *et al.* (2000). They found dissociable neural mechanisms underlying the implementation of cognitive control and performance monitoring, suggesting that these two psychological processes are separable. The authors used a variant of the Stroop task in order to study these phenomena (Stroop, 1935). Participants were given an instruction either to read the word or to name the color of an upcoming Stroop stimulus. Thus, they had an opportunity to prepare for the task at hand. Stimuli could be congruent (in which the word and its ink color matched, such as the word 'green' printed in green ink) or incongruent (in which the word and its color mismatched, such as the word 'green' printed in red ink). On the one hand, the authors found that left dorsolateral prefrontal cortex was more active when subjects were instructed to name the color rather than read the word, evidence that this region was implementing cognitive control by preparing for the more challenging task. On the other hand, the anterior cingulate cortex was activated by the stimuli itself, showing greater activation for the high-conflict incongruent compared with the low-conflict congruent stimuli. This pattern of dissociation enabled a hypothesis that the lateral prefrontal cortex was implementing cognitive control while the anterior cingulate cortex was involved in the monitoring of performance.

DOCUMENTING INDIVIDUAL DIFFERENCES

The first three approaches we described above, all rest on a *consistent* mapping of brain and behavior that is found across individuals. However, there have been recent lines of research that extend the use of fMRI to the identification of differences across individuals as well. Consider, for example, a study by Canli *et al.* (2002). They found consistent activation in the amygdala among participants when viewing fearful facial expressions, but inconsistent activation when participants viewed happy facial expressions. A good deal of the variability in activation when viewing happy expressions was predicted by measuring participants' scores on an extraversion scale. The higher the score on this scale, the higher was the activation in amygdala. The authors discuss the difference between the responses to fearful and happy expressions in terms of the adaptive value in responding consistently to fear, in contrast to the tailored responsiveness of outgoing people to happiness.

Using similar methodology, Schwartz *et al.* (2003) found that adults who were identified as inhibited toddlers showed higher activity in the amygdala for novel *vs* familiar faces compared with adults who were identified as uninhibited toddlers. One possibility raised by these data is that the difference in temperament exhibited by inhibited and

**NOT FOR
PUBLIC RELEASE**

uninhibited toddlers may be the result of differential activity in the amygdala in response to novelty, and that neural properties relating to temperament may be preserved from childhood into adulthood (Schwartz *et al.*, 2003).

These and other studies of individual differences in brain activation can play a role with behavioral data in accounting for both consistent and inconsistent behavior across tasks.

TESTING PSYCHOLOGICAL MODELS

As we come to understand more and more about the functionality of regions of the brain, it is becoming increasingly possible to use fMRI data to distinguish alternative psychological models of task performance. For instance, Brown and Braver (2005) used fMRI to distinguish between two competing theories of cognitive control. One theory posits that in situations where multiple responses compete for behavior, a conflict monitor calls for increased control to mitigate interfering responses (Botvinick *et al.*, 2001). Supporters of this theory have demonstrated that the anterior cingulate cortex appears to be responsive to the degree of response conflict in a variety of cognitive tasks and that interactions between the anterior cingulate and lateral prefrontal cortex correlate with behavioral adjustments in situations of high conflict (Kerns *et al.*, 2004). An alternative account is that the anterior cingulate responds to the likelihood that an error will follow from a given context. This theory builds upon the reinforcement learning literature, which demonstrates that a dopaminergic error signal drives learning, allowing an organism to adapt its behaviors to specific contexts. By carefully dissociating error likelihood and conflict through a change-signal task, Brown and Braver (2005) found that the anterior cingulate learns to respond to situations where errors are probable even when conflict is low. These results are consonant with an error likelihood account of cognitive control rather than a conflict monitoring theory. Importantly, the authors point out that both the error likelihood model and the conflict monitor model accurately fit behavioral data from the change-signal task, and that it was only the fMRI data that differentiated these two models of cognitive control (Brown and Braver, 2005). So we see here that fMRI can be used to test model predictions and thereby distinguish among competing psychological theories even when behavioral data alone may not make that discrimination.

CONCLUSION

Of what value have fMRI data been to the study of mind and brain? Let us count the ways. They have been instrumental in establishing correlations between brain and behavior. They have allowed us to examine overlapping and non-overlapping patterns of brain activation that are valuable in building up a view of shared and distinct processes among psychological tasks. They are beginning to permit us

to understand consistencies and inconsistencies in human behavior, as accounted for by consistencies and inconsistencies in brain activation. Finally, they are now allowing us to test among alternative psychological models of behavior. We are not advocating a neuroimaging-imperialism here because it is quite clear that the true value of neuroimaging data comes in concert with sophisticated behavioral data collected from normal and brain-injured participants. Nevertheless, there have been substantial accomplishments with fMRI data in fewer than 20 years.

REFERENCES

- Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–52.
- Brown, J.W., Braver, T.S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, 307, 1118–21.
- Canli, T., Sivers, H., Whitfield, S.L., Gotlib, I.H., Gabrieli, J.D.E. (2002). Amygdala response to happy faces as a function of extraversion. *Science*, 296, 2191.
- Coltheart, M. (2006). Perhaps functional neuroimaging has not told us anything about the mind (so far). *Cortex*, 42, 422–7.
- Downing, P.E., Jiang, Y.H., Shuman, M., Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293, 2470–3.
- Eisenberger, N.I., Lieberman, M.D., Williams, K.D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, 302, 290–2.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–30.
- Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. *Trends in Cognitive Sciences*, 10, 64–9.
- Jonides, J., Nee, D.E., Berman, M.G. (2006). What has functional neuroimaging told us about the mind? So many examples, so little space. *Cortex*, 42, 414–7.
- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3, 759–3.
- Kerns, J.G., Cohen, J.D., MacDonald, A.W., Cho, R.Y., Stenger, V.A., Carter, C.S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, 303, 1023–6.
- MacDonald, A.W., Cohen, J.D., Stenger, V.A., Carter, C.S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, 288, 1835–8.
- Poldrack, R.A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10, 59–63.
- Schwartz, C.E., Wright, C.I., Shin, L.M., Kagan, J., Rauch, S.L. (2003). Inhibited and uninhibited infants “grown up”: Adult amygdalar response to novelty. *Science*, 300, 1952–3.
- Shelton, J.R., Fouch, E., Caramazza, A. (1998). The selective sparing of body part knowledge: a case study. *Neurocase*, 4, 339–51.
- Smith, E.E., Jonides, J. (1995). Working memory in humans: neuropsychological evidence. In: Gazzaniga, M., editor. *The Cognitive Neurosciences*. Cambridge, MA: MIT Press, pp. 1009–20.
- Smith, E.E., Jonides, J., Koeppe, R.A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex*, 6, 11–20.
- Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 12, 643–62.
- Wager, T.D., Rilling, J.K., Smith, E.E., et al. (2004). Placebo-induced changes in fMRI in the anticipation and experience of pain. *Science*, 303, 1162–7.