

18 Neuroimaging as a Tool for Functionally Decomposing Cognitive Processes

William Bechtel and Richard C. Richardson

Both those who extol and those who castigate neuroimaging studies and their invocation in cognitive science often misconstrue the contribution neuroimaging is seeking to make, and is capable of making, to cognitive science. We do believe that advances in neuroimaging, including functional magnetic resonance imaging (fMRI), as well as such techniques as single cell recording, are important contributions to the experimental repertoire of cognitive science. We also anticipate that neuroimaging's importance will increase with improvements in imaging technologies and techniques.

Our objective here, however, is neither to extol nor to castigate neuroimaging, but to make clear what sort of contribution neuroimaging, when done well, can make to understanding and explaining mental phenomena. Many of those who adopt extreme views of neuroimaging are not themselves practitioners of the technique and fail to appreciate principles that are commonplace or platitudes among the expert practitioners. Of course, some who adopt extreme views are practitioners. Accordingly, though we are by no means expert practitioners, we start with some commonplaces that are not always transparent in reports of neuroimaging, but are generally understood by the researchers conducting the research (see, for example, Logothetis 2008) and need to be taken into account by those evaluating it:

1. The brain contains some regions that are specialized for processing specific types of information. This is most clearly established in sensory and motor regions, though we expect the conclusion is more general; yet even these regions integrate information from a large number of other regions with which they engage in complex dynamical interaction. Some of those regions are colloquially thought of as "downstream" in the visual system.
2. Functional MRI, by identifying particular areas of increased blood flow, may seem to support the idea that there are highly specialized regions, some kind of modules, but much of this is illusory. Measures of BOLD activity show differences in regional activation, and the impression of modules is enhanced by the often vivid illustrations. Finding differences in activation, coupled with BOLD signals, is not tantamount to

identifying modules; and the typical and very vivid coloring is but an illustrative artifact.

3. Even though there are specialized processing regions (commonplaces 1 and 2), these regions are not encapsulated or insulated from one another. A given region may process a particular type of information and thereby be differentiated from other regions, but it processes information generated in other areas with which it is connected and outputs information to yet other regions. This is often accompanied by elaborate feedback (even in the sensory areas). The network of connections is crucial to the operation of the brain (commonplace 7). Though there is often talk of centers, the real interest is in seeing how various areas interact in a given task. What is wanted is an understanding of populations of neurons, their connections, and ultimately their dynamics.

4. The fact that a given brain region displays increased activation to a stimulus or in performing a specified task does not at all imply that the same region does not respond to many other stimuli or cannot engage in other tasks (commonplace 3). In fact, aside from more peripheral areas, regions commonly respond to many different stimuli; there is also a great deal of spontaneous activity. This is one implication of emphasizing that BOLD signals tell us about *changes* in activation.

5. Lack of a BOLD signal in a region doesn't mean the region lacks activation. This is true for a number of reasons: (a) Though it is true that enhanced activation increases blood flow, it is possible for functionally significant regions to go undetected (a voxel includes thousands of neurons, and it takes increased activity in many to "light up" a region); (b) some neurons may be more efficient than others, and thus have a less visible BOLD signal; (c) some regions have very high blood flow in general, and their BOLD signal may increase only marginally even when active.

6. The activity in any given region of the brain is a function of both excitation and inhibition; any changes in the balance—whether the result is an increase or a decrease in activity—will affect the BOLD signal. (Increased inhibition can lead to increased metabolism, and not to a negative signal.)

7. Imaging studies are invoked not primarily to answer the question of where in the brain an operation is performed, but to determine what operations are performed and how they matter to the overall outcome. Moreover, they are not intended to address global issues (e.g., what sort of information processing system the brain is) but local issues (e.g., whether two activities invoke the same neural processes, or whether some area is implicated in some cognitive activity).

We take these to be uncontroversial ideas among those whose research deploys fMRI systematically. We also do not take these to be unconnected observations. Indeed, they form a more or less integrated set of practices. Unfortunately, some interpretations of fMRI work defy one or more of the commonplace assumptions within the field.

We emphasize these commonplaces at the outset because we think it is important to keep the overall framework in mind when discussing the use and abuse of fMRI. We

will return to specific commonplaces in our discussion, which focuses on Loosemore and Harley's challenges to invoking neuroimaging results that engage cognitive phenomena. Loosemore and Harley's challenges are representative of those who castigate neuroimaging as useless for understanding the mind/brain (for other recent critics, see Coltheart 2004; Uttal 2001; van Orden and Paap 1997). Their critique has two major parts: (1) an argument that neuroimaging is premature, since its results can only be properly construed once cognitive psychology has reached a more mature state; and (2) an analysis of six studies (four involving neuroimaging, one single-cell recording, one transcranial magnetic stimulation) that are meant to demonstrate the inability of neuroimaging to contribute to the maturation of cognitive theories. In the following sections we address these points. In the second section, we contrast Loosemore and Harley's conception of how sciences progress with a different conception grounded on analysis of progress in sciences devoted to discovering the mechanisms responsible for phenomena of interest. In the third section, we examine the neuroimaging studies Loosemore and Harley criticize and argue that the objections the researchers raise against them are not warranted.

The Contribution of Neuroimaging to Discovering Cognitive Mechanisms

Loosemore and Harley offer what we take to be an extremely problematic account of how scientific inquiry progresses as it attempts to identify the mechanisms responsible for a given phenomenon. They present three stages cognitive science is "destined" to go through in its history. Stage 1 involves advancing "metaphor-like descriptions of functional-level mechanisms." At stage 2, researchers agree on a "complete outline theory of the human cognitive system." At this stage, Loosemore and Harley further claim, "we would expect the basic processes and structures to be clear enough that no drastic changes would be arriving to disrupt the outline theory in the future." Finally, at stage 3 researchers can identify how this complete account is realized in the human brain.

The talk of the stages of cognitive science as a matter of destiny is theirs rather than ours, though it is not developed in their paper to any significant extent. Apparently, the stages are inflexible in ordering, and determine the usefulness of information from the neurosciences. We regard their stage theory with skepticism. Loosemore and Harley maintain that cognitive science is currently only at stage 1 and until it progresses through stage 2, neither neuroimaging nor any other investigation of the brain is capable of providing any significant contributions. In this view, the role for neuroimaging would essentially be limited to identifying the structures that realize a given cognitive process; neuroimaging would not contribute substantially toward elaborating the structure of cognitive models, and would be useful only at stage 3. (Presumably, other sources of information about neurophysiological mechanisms would also not contribute, in the absence of at least a stage 2 psychology.) In our view, this both

underestimates the contributions neuroimaging can make to the development of cognitive models and overestimates the prospects of progress for cognitive science in the absence of information about neural processing. Integration of behavioral, psychological and neuroscientific perspectives is, on our view, much more promising as a research program (Bechtel and McCauley 1999; Craver 2007; McCauley 2007; McCauley and Bechtel 2001; Richardson 2009). It would be a mistake to postpone inquiries into cognitive neuroscience while we wait on the emergence of a completed psychology; moreover, it is possible for the use of imaging data to discriminate between cognitive models (Henson 2006).

Figuring out the basic processes or operations through which a complex mechanism is able to generate some phenomenon of interest—in this case, some aspect of human behavior or some cognitive capacity—is, in fact, a critical part of developing a mechanistic explanation. Guidance in discovering these operations often comes from identifying the component parts of the system and using their identity as a tool in identifying the operations they perform. Cognitive psychology, through much of its history (from the 1950s through the late 1980s), was constrained to attempting to identify the operations involved in producing cognitive phenomena without benefit of information about the brain regions that perform these operations.¹ This was a result of the lack of research tools and techniques that enable identification of human brain areas involved in performing cognitive activities.

During the same period, researchers in neuroscience developed not only detailed accounts of the brain areas involved in processes such as visual perception, but also elaborated serious proposals as to the specialized operations (commonplace 1) these areas perform (van Essen and Gallant 1994). They did this working with other species in which invasive techniques such as single-cell recording, direct electrical stimulation, and localized lesions could be employed (Logothetis and Pfeuffer 2004). It is important to realize that these neuroscientific inquiries were concerned with determining the identity of brain areas (using tools such as neural connections and the patterns of topographical projection of the visual field onto different maps), systematically mapping out the projections to and from a given area (using, for example, retrograde stains to detect regions projecting to an area of interest), and figuring out what operations the delineated areas performed. The last objective was accomplished primarily by determining the nature of the visual stimuli that would specifically produce action potentials in neurons in a particular brain region (edges in V1, illusory contours in V2, motion in V4, color and shape in V5, etc.). The research produced some quite surprising discoveries about the functional processes involved in vision, including the differentiation of two relatively independent processing streams (dorsal and ventral) that process different information about visual stimuli—the identity of the objects perceived and the location and potential for acting on the objects (see Bechtel 2008, for a historical review).

The results of these investigations, mostly done on monkeys (also mice, rats, and rabbits), have been generalized to humans, although often only via the use of neuroimaging since research comparable to that done on other mammals cannot be performed on humans (for ethical and political reasons). Arguably, these accounts of visual processing generated in approximately fifty years of research are the most detailed characterizations of the operations involved in a cognitive activity (commonplace 1). Moreover, these results are among those least likely to be disrupted (radically at least) by further research. This was due in no small measure to the fact that research identifying brain regions went hand-in-hand with research identifying the operations these regions performed: As researchers differentiated brain areas, for example, on the basis of their containing a topological map of the visual field, they also investigated what stimuli would enhance the spiking rate of neurons in the area. The elaboration of this research depended on coordinating research from a variety of scientific research specialties. It was, emphatically, not true that psychologists first developed a detailed and defensible cognitive model, which was only afterward shown to be realized in the brain.

Stepping back a bit, we can place research seeking mechanisms for cognitive phenomena in the broader context of the discovery of mechanisms in the life sciences (Bechtel 2008; Bechtel and Richardson 1993). Developing a mechanistic explanation requires, in part, decomposing the mechanism into its parts and operations. Another extremely important challenge in developing a mechanistic explanation is to determine how the parts and operations are organized, both spatially and temporally, so as to produce the phenomenon. This is crucial to understanding the discovery of mechanisms, but perhaps not so crucial in the current context, except in the emphasis on uncovering system dynamics.

Typically, different research techniques are required to decompose a system functionally into its operations or structurally into its parts, and one set of tools may be available when another is not. Brodmann (1909; 1994), for example, made major progress in differentiating brain regions using neuroanatomical tools. Despite hoping the areas he delineated would be functionally significant, he had no tools for identifying the operations they performed. Likewise, gene sequencing has provided detailed maps of the genomes of many species, but cannot itself inform us about the functionality of various strands of DNA. On the other hand, decomposing the mechanism functionally into its operations can also be pursued without information about the structures involved. Like cognitive psychologists, biochemists initially had no tools for discovering the structures they took to be responsible for catalyzing reactions and had to proceed to identify the reactions employing purely functional techniques. Whereas cognitive psychologists have had limited tools for developing functional decompositions (e.g., dissociations in reactions times, error patterns, or cognitive deficits), biochemists were able to be more invasive by, for example, inhibiting reactions with poisons and identi-

fy the product that built up. Nonetheless, the fundamental logic of the cases is similar.

The fact that researchers have made progress in developing a functional decomposition independently of a structural decomposition, or vice versa, however, does not make it a virtue to proceed with one before the other. Though psychology did manage to make progress in the middle decades of the last century in identifying cognitive processes, the virtue was born of necessity. Without access to an independent source of evidence, psychologists made the best of what was possible. Proposing that operations are performed by identifiable parts often serves to significantly advance inquiry into both parts and operations. Typically, neither the functional analysis nor the structural analysis reaches a mature form before operations are identified with parts; neither can it reach a mature form the organization of the system is understood. It is not uncommon for researchers to realize, after they have linked an operation with a part, that the part in question cannot perform the whole operation and, in fact, multiple operations are involved in what seemed to be a simple operation. In some cases, knowledge of the structure and capacities of a part suggest that the characterization of the operation itself needed to be revised. For example, biochemists spent 20 years seeking a high-energy chemical intermediate that transferred energy released in the electron transport chain to adenosine triphosphate (ATP). The recognition that this reaction occurs in the cristae membranes in mitochondria suggested a very different operation for energy transfer, one involving the creation of a proton gradient across a membrane. Likewise, recognition of the importance of long-term potentiation (LTP) in the 1960s involved both anatomical work and electrophysiological research integrating both structural and functional information (Craver 2003).

Linking operations to parts is thus often an important aspect of developing both the functional and structural decompositions (commonplace 1). Since the structural decomposition often involves specifying spatially where the parts are, and the connections among those parts, integrating the two often takes the form of localizing operations within the overall mechanism (accordingly, we employ the term *localization* for hypotheses identifying structures with functions). The point of localization, however, is not simply to figure out *where* operations take place, but to draw on what is known about the structures involved and their relation to other structures to provide evidence and often heuristic guidance in characterizing operations and unearthing the systemic organization. As we emphasize in our commonplaces (3 and 7), the point is to understand how the system functions, how it works dynamically, and that is not just a matter of finding what, if anything, realizes a particular function. Thus, one should not construe fMRI, or other means of localizing cognitive processes in the brain, as simply revealing how an independently developed psychological theory is implemented, or where some cognitive capacity is realized.²

The process of developing a mature mechanistic account may involve numerous iterations of localization combined with further articulations of functional and struc-

tural decompositions, along with independent assessments of the functional capacities of components. For example, initial attempts to identify where in the brain a supposed operation involved in a particular task occurs may reveal multiple structures; it may also fail to reveal some structures that are involved (commonplace 5). This is a powerful spur to discovery. It is possible that several of the brain areas perform the same operation. There is sometimes substantial redundancy. It is also possible that each of the brain areas performs a different function, which complement one another.³ Typically, researchers tend to suspect that each distinct part is performing a somewhat different operation and that the initial functional account needs to be amended. Since invasive research on brain areas in humans is restricted, researchers need to pursue other strategies for revealing the operations involved. An important clue is often the discovery that one of the brain areas is involved in a variety of different tasks (commonplace 4). Researchers can then explore what operation might be employed in these different tasks, what different tasks have in common, and how differences can be managed. Having connected one of the areas to a specific operation, they can then return to the first task and ask what operations are required besides that associated with the given area. The process now iterates as researchers explore these other areas. The important point to emphasize is that localization need not come at the end of inquiry (stage 3 in Loosemore and Harley's picture), but may come early in inquiry, where it can function as a discovery heuristic (Wimsatt 2007). Another way to press the point is to emphasize that functional decompositions both respond to and shape structural decompositions.

Although Loosemore and Harley emphasize the stage account for developing cognitive psychology in their abstract, it is not totally clear to us how committed they really are to it. As we have said, we are skeptical. Nothing much here hinges on the skepticism over stages. The paper included here does not do much to elaborate the account. Instead, they articulate four levels at which "imaging the localization of function of components of anything—parts of cars, computers, or brains—can be described": tokenism, absolute location, relative location, and functional. Tokenism is the use of brain images just because they are available and impressive, not because they contribute to the scientific investigation. Of course, we recognize that sometimes the imaging doesn't substantially contribute to research, aside from providing a vivid presentation. We suppose that is tokenism. Absolute location provides the locus of an operation, whereas relative localization appeals to how operations are situated with respect to each other to support claims about how they act on each other. Loosemore and Harley note that, in the case of the brain, relative location can be further articulated in terms of neural connectivity, since it is connected brain areas that are likely to interact with each other in generating the phenomenon. The functional location uses the location where an operation is performed to develop the decomposition into operations—the contribution of a component to systemic function. As soon as they introduce this level, Loosemore and Harley return to their central criticism: "you already need to

know mostly what each component does before you can begin to make sense of it.” This rests on the assumption that the only real use of neuroimaging would be to provide information about the implementation or realization of already well-understood cognitive processing. We regard this as their central mistake.

If this charge were true, it would not be necessary to examine actual imaging studies to show that, although they might seem to be at level 4, they really are only at level 2 because they do not contribute to the further articulation of a functional decomposition. All that would be necessary would be to show that a detailed functional decomposition (that was unlikely to be further changed) was not already available in cognitive psychology to demonstrate that imaging studies were limited to stage 2 or 3. The most Loosemore and Harley would need to show is that the psychology is so impoverished that it cannot sustain any attempt at neural realization. The critical focus would be on the psychology rather than the neuroscience. As an analogy, absent a developed phenomenological theory of inheritance (such as the patterns supposedly unearthed by Mendel), it might make little sense to inquire into the mechanisms of inheritance (though that was actually attempted). The actual history is more interesting, and surely does not sustain the top-down picture that dominates in Loosemore and Harley.

In analyzing the cases, they argue that the imaging studies have not contributed toward a better functional decomposition. Thus, implicitly at least, Loosemore and Harley seem to be granting that imaging studies could contribute to functional decomposition, as we have suggested, but in fact they fail to fulfill this promise. To address this, we need to go beyond a general characterization of a mechanistic research program and the role localization plays in it to a detailed examination of the contribution of localization arising from neuroimaging.

Reexamination of the Neuroimaging Studies

A major portion of Loosemore and Harley’s discussion is devoted to six target neuroimaging studies (two, in fact, do not employ imaging but either single-cell recording or transcranial magnetic stimulation to suppress operations). They choose the studies, they explain, on the basis that they were discussed in the popular press—a selection strategy they defend on the grounds that they “wanted to generate a sample of what the popular press finds interesting about psychology and the brain.” This strikes us as a peculiar way of identifying neuroimaging studies that have the best chance of contributing to a functional decomposition of cognitive processes. Even if, as Loosemore and Harley claim, the press is interested in research that purports to explain behavior, it might not be the best evaluator of what research is contributing positively to current research objectives any more than the popularity of a topic is indicative of its importance (think of flying saucers or Bigfoot). It is also clear that reports in the public press tend to distort the studies at issue, either emphasizing what the studies do not, or

actively distorting their impact. The best research is more complex, and more focused, than the press is comfortable reporting. For those who have any doubt, the interest of the press in claims of finding genes for particular functions and underlying specific diseases, rather than in the complex regulatory mechanisms involved in gene expression that are increasingly being identified, should be a cautionary note. We regard the popular reports of brain regions “for” this or that cognitive activity with comparable skepticism. Irrespective of how these studies were chosen, however, we will argue that each aims at contributing to functional decomposition of mental processes and so escapes the charges Loosemore and Harley make.

Before turning to the studies themselves, we need to comment briefly on a peculiar feature of Loosemore and Harley’s critique of these studies. Part of their critiques is directed to the details of the studies themselves, and for the most part, we’ll be occupied with this part. However, a second part of their critiques is devoted to arguing that an alternative framework they put forward, employing a very different functional decomposition, is compatible with the results of each of the studies. As a result, they claim that the functional decompositions of cognitive phenomena are underdetermined by the evidence and any imputed functional contributions of the imaging studies are unsupported. We would be remiss to ignore this second part.

Loosemore and Harley present their framework, which they call a *molecular model of cognition*, as “a new, unified framework that describes the overall architecture of the human cognitive system.” They evidently do not think of this as a serious model of cognition, but as a kind of toy structure they can use critically. The architecture consists of a dynamic working memory in which active “concept units” representing instances interact with each other based on constraints incorporated in each concept, and a background long-term memory in which concept units representing generic categories reside. The model is presented very sketchily, with no empirical backing and no substantive constraints. Loosemore and Harley describe qualitatively how their model might be applied to recognizing visual objects and the activity of sitting down, but offer no detailed results to demonstrate that the model could accommodate even the most accepted empirical results about recognition or control. From Loosemore and Harley’s perspective, this is not a serious problem, since their goal is simply to show that the neuroimaging results underdetermine the functional decomposition of cognitive mechanisms. As they explain, “our real goal is to see how sensitive the conclusions of these studies might be to a slight change in the theoretical mechanisms whose location is being sought.” One might think that if research results can be equally accommodated by newly constructed and untested qualitative models, this is a reason not to trust the imaging researchers’ interpretation of their results. This, however, is a cartoon of serious science. Scientists do not need to defend the interpretation of their research against any underspecified model, but only those that are taken to be serious alternative models. (If Loosemore and Harley were right, then evolutionary biologists would

need consistently to defend themselves against Creationist contentions. That would hardly improve the state of the biological sciences.) The alternative models that are taken seriously are those for which evidence already points to their plausibility. In the absence of any articulated models, perhaps speculation is sufficient. Normally, though, there is a more or less well-articulated set of alternatives, and the point of experimentation is to discriminate among them (commonplace 7). We'll see that this is a pattern Loosemore and Harley miss in some of the studies they condemn.

Loosemore and Harley indicate that their alternative cognitive model is in the tradition of computational models in cognitive science. Serious computational modelers, however, provide detailed accounts of operations and demonstrate that the resulting models, minimally, can accommodate behavioral data about the phenomena being modeled. There is increasing interest in computational models linking their results to knowledge of the brain, often based on neuroimaging results (see, for example, Anderson 2007). These efforts, however, are highly demanding, and importantly, when they attempt to draw on neuroimaging studies to evaluate their models, the project is essentially comparable to that in the imaging studies Loosemore and Harley review. Our point is not to defend this sort of modeling. Computational models are sometimes dramatically underdetermined by the evidence, so that many computational models are compatible with both behavioral profiles and capabilities. This renders the endeavor extremely challenging. For those engaged in such research, the goal is to reduce the degree of underdetermination over time as new evidence eliminates previously plausible models. This underdetermination is different from that resulting from an underspecified and untested model such as Loosemore and Harley invoke.

Before leaving Loosemore and Harley's molecular model, it is worth noting that one of its more novel features is the proposal that the representations of instances are "created on the fly." There are interesting similarities here with Larry Barsalou's account of concepts, but the differences in how Barsalou has developed and defended his account are also noteworthy. In his early research on concepts Barsalou (1987) demonstrated, using behavioral data, that prototypicality judgments for both goal-directed and ordinary taxonomic concepts vary substantially over time, and proposed that they were constructed anew and differently on each occasion of use. Barsalou (1999) advanced an account according to which concepts are grounded in perceptual-motor processing. Accordingly, unlike many traditional psychological accounts of concepts that treat them as amodal representations, Barsalou argues that concepts are essentially modal. More recently, he has defended a view that reasoning with concepts involves simulation, where simulation is understood as

... the reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind. As an experience occurs (e.g., easing into a chair), the brain captures states across the modalities and integrates them with a multimodal representation stored in memory (e.g., how a chair looks and feels, the action of sitting, introspections of comfort and

relaxation). Later, when knowledge is needed to represent a category (e.g., chair), multimodal representations captured during experiences with its instances are reactivated to simulate how the brain represented perception, action, and introspection associated with it. (Barsalou 2008, 618–619)

Unlike Loosemore and Harley, who contend that the full functional account of these processes needs to be developed *in advance* of finding neural realizers, Barsalou has embraced neural studies, including imaging, as a tool for developing his account. In his hands, fMRI results are employed in much the same manner as purely behavioral evidence in determining the linkages between the operations involved in conceptual tasks and those figuring in sensory motor tasks. For example, in a purely behavioral study that supported the claim that simulation figures in perceptual processing, Solomon and Barsalou (2004) took advantage of the fact that larger properties are more difficult to verify perceptually to predict that it should be more difficult to evaluate whether larger properties apply to a given concept (a property verification task) than smaller properties. This is a prediction an amodal account of concepts would not make, but turned out to be true. Similarly, Pecher, Zeelenberg, and Barsalou (2004) found that switching modalities in successful property identification trials impaired performance. Simmons et al. (2007) turned to fMRI to show that areas in left fusiform gyrus that are active in color perception are more active in tasks requiring subjects to verify that properties are related to a concept when subjects are queried with color properties than with motor properties. These are results that would be expected if concepts are modally grounded, but not otherwise. Note that Barsalou had little interest in the specific brain areas involved but rather was interested in whether the same brain areas are involved in sensory processing and categorization tasks, as this could contribute to characterizing the operations involved in the categorization task. Again, this is a commonplace (3).

We turn now to the specific studies Loosemore and Harley review and criticize. Of course, criticism of such studies is welcome and often informative. Studies are sometimes well constructed and sometimes not. We had no prior conviction about whether or not the studies Loosemore and Harley criticized are well constructed. In fact, we find some more interesting than others and have ordered the four neuroimaging studies in ascending order of interest (taking up the two nonimaging studies last). Our concern, though, is with whether the sort of criticism Loosemore and Harley offer is probative—whether it sheds light on the use or abuse of neuroimaging. Let us look at the cases Loosemore and Harley selected.

Dux, Ivanoff, Asplund, and Marois

The first study discussed by Loosemore and Harley is one by Dux, Ivanoff, Asplund and Marois (2006). It has long been known that tasks may compete for cognitive resources, and when interference occurs, tasks appear to be queued rather than performed in par-

allel. Behaviorally, this was explored with studies of reaction time. What was generally seen is that competing tasks tended to be delayed if they were presented within a suitably short timeframe. The delay was both a measure of the interference and suggestive of a mechanism. The question Dux et al. ask concerns the neural basis of this interference. Until their study, attempts to localize the “bottleneck” had relied on the amplitude of BOLD responses to identify the source of the interference. The result is that different studies have emphasized a wide array of “putative neural substrates,” including lateral, frontal, prefrontal and dorsal regions. The novel contribution of Dux et al. was to turn to time-resolved fMRI, allowing them to discern the relative timing across regions. They argue that the posterior lateral prefrontal cortex (pLPFC) is crucially implicated in the limitations on processing. They do not claim that it is solely implicated in the limitations on processing.

Loosemore and Harley say this is a level-2 or level-3 study, telling us the absolute location of some process, and perhaps something about its relation to other psychological processes. Loosemore and Harley express some skepticism about whether bottlenecks are real or not. Of course, if they are not, then the Dux et al. study can hardly reach even level 2, since there would be no process to localize. We are not certain what weight to give this skepticism.⁴ The existence of interference is a robustly resilient psychological effect, quite apart from the use of imaging studies. We are inclined to take the allusion to a “bottleneck” as simply shorthand for the interference effects evident in competing tasks. At other places, Loosemore and Harley seem to relax their skepticism, at least acknowledging that there are interference effects. They notice that “cognitive psychological studies alone” are sufficient to underscore the fact that there is interference. They also note that the Dux et al. study does not allow the researchers to “discover the ‘reason’ people find it hard to do two tasks at once.” With that conclusion we have no quarrel. Dux et al. also do not claim to have found such a “reason.”

As we’ve said, we take the point of the Dux et al. study to adjudicate between the several proposals for the locus of the “bottleneck,” using time-resolved fMRI. The pLPFC exhibited behavior consistent with it being a “bottleneck.” A bottleneck needs to satisfy several criteria: (1) it must be shared by tasks that do not share either input or output modalities; (2) it must be “involved” in response selection; and (3) it must exhibit serial queuing. The third is, of course, what makes a bottleneck a bottleneck. It is important that being a bottleneck at all depends on the structure being connected in appropriate ways to both input and output modalities. If there were not multiple inputs, there would likewise be no possibility for being a bottleneck (criteria 1 and 2). This is not a matter so much of finding a location for some process as of finding an appropriately linked brain structure (commonplace 7). Dux et al. cautiously observe that this does not mean no other regions play a role in dual task interference (commonplace 4); it is simply *not* true that the failure to detect activation in their experiment indicates there is no activation in the alternative regions, or that other regions

are not functionally involved (commonplace 2). As they notice, in addition, this region is recruited by quite diverse cognitive tasks. Finally, the Dux et al. study doesn't actually pretend, as far as we can see, to tell us anything about the specific cognitive function(s) in the pLPFC (commonplace 7) aside from the fact that it is, in one way or another, involved in response selection (criterion 2). It is nonetheless a functional study, geared to identifying the network of connections involved in performance.

Aron, Fisher, Mashek, Strong, and Brown

For Aron, Fisher, Mashek, Strong, and Brown (2005), the goal is to understand the type of mental operations involved in early stages of romantic love, which would seem to situate their study at level 4. The main alternatives they consider are that romantic love is a distinct emotion (as has been defended by Gonzaga et al., 2001) or that it involves operations that figure more generally in the pursuit of rewards. Loosemore and Harley mischaracterize these alternatives as "a strong emotion" or "an overwhelming desire to achieve an objective," missing the point of relating romantic love to the broader set of processes involved in pursuing rewards. To evaluate these alternatives Aron et al. consider two predictions that would follow from the second alternative but not from the first: (1) "romantic love would specifically involve subcortical regions that mediate reward, such as the ventral tegmental area (VTA) and ventral striatum/nucleus accumbens," and (2) "the neural systems involved in early-stage romantic love ... would be associated with other goal and reward systems, such as the anterior caudate nucleus." The imaging results they present support these predictions. Note how Aron et al. state their conclusion:

[T]he results lead us to suggest that early-stage, intense romantic love is associated with reward and goal representation regions, and that rather than being a specific emotion, romantic love is better characterized as a motivation or goal-oriented state that *leads to* various specific emotions such as euphoria or anxiety. (p. 335)

Thus, Aron et al. present their study as identifying parts of a network of processes involved in romantic love, not, as Loosemore and Harley suggest, as equating romantic love with a particular process (commonplace 7). Notice, the thought is not that there are not specialized regions (commonplace 1), but that romantic love involves the VTA, which "mediates" reward. The thought is also not that the VTA is a center that "realizes" reward mechanisms. In fact the study is focused on *denying* there is a distinct emotion, and situating the response in a more general network of responses.

One of the reasons Aron et al. find it significant that the VTA and related areas are involved is that these areas are dopamine rich. Loosemore and Harley criticize this part of the study for not specifying the relation between dopamine release and motivation. However, that was not Aron et al.'s objective. Rather, the point of identifying these areas as dopamine rich is to be able to link the imaging results with a broad array

of results from other techniques for investigating dopamine function and to relate romantic love to other activities that involve these dopamine-rich brain areas. Loosemore and Harley question whether the study shows anything more than that early stages of romantic love involve desire for the object of love. What the study actually purports to show is that the desire involved in romantic love involves the same mechanisms of desire as figure in such phenomena as desire for cocaine and so is not unique to romantic love. If these results are correct, they advance the functional study of romantic love by relating it to a broader class of mental activities and the network involved in those activities. Of course, much more research is required to identify the component parts and operations within this network, and here we are skeptical; but determining what network to examine is an important step in developing a functional account. Our point is that the focus of the study is on networks rather than loci (commonplace 7).

Bahrami, Lavie, and Rees

Bahrami, Lavie, and Rees (2007) addressed the question of whether modulating attentional load in one task could affect the processing in early visual areas of stimuli of which the subjects lacked conscious awareness. The study is clearly about the relation of one process (attention) to another (visual processing). This is made apparent by the motivation for the study. Bahrami et al. turned to imaging to address a possible confound in earlier attempts to address this issue behaviorally by measuring priming effects through reaction times (Chun and Jiang 1998; Dehaene et al. 1998). The effects of priming on reaction time could result from the effects of attention on early processing or from modulating the strength of motor response associations. The latter option was not an idle worry, as there is independent psychological evidence for the effects of priming on motor responses (Sumner et al. 2006). Also, the previous evidence that high perceptual load modulates V1 activity did not distinguish between conscious and unconscious perception: "Indeed, some have even claimed that V1 activity related to feedback from extrastriate cortices serves as the arbiter of conscious awareness" (Bahrami et al. cite Silvanto et al. 2005). Given the effects they were able to show on processing in V1 under conditions where they prevented subjects from becoming aware of the visual stimuli while otherwise modulating their attentional load, Bahrami et al. concluded the following:

The present findings are the first to show that neural processes involved in the retinotopic registration of stimulus presence in V1 depend on availability of attentional capacity, even when they do not invoke any conscious experience. . . . Importantly, our new findings that the level of attentional load in a central task determines retinotopic V1 responses to invisible stimuli clarify both that unconscious processing depends on attentional capacity (which is reduced in conditions of high load) and that availability of attentional capacity for stimulus processing (in the low load conditions) cannot be a sufficient condition for awareness. (p. 510)

It is hard not to see this study as directed at level 4, the functional level in Loosemore and Harley's hierarchy. The objective was to find out whether one process (attentional demands in one task) affected another (early visual processing). The only reason Loosemore and Harley offer for it not being at this level is that their molecular framework could offer a different interpretation. At best, this would show that the Bahrami et al. study was not conclusive, not that it was not a level-4 study. Since they later claim that none of the studies they review rise above level 2, there must be another reason for downgrading the study. Our best guess is that Loosemore and Harley infer from the fact that Bahrami et al. invoke processing in a place in the brain that it is only a localization study, rather than a contribution to a functional analysis (common-place 7). But it is clear that location of processing was used here only to situate the activity of interest in an already developed functional model of visual processes so as to gain evidence about the impact of attention on that processing.

Haynes, Sakai, Rees, Gilbert, Frith, and Passingham

The goal of Haynes et al. (2007) is likewise to resolve a functional question, one that had been posed by earlier imaging studies revealing prefrontal activities in human goal-related activities. These studies failed to resolve whether the activity in prefrontal cortex reflected an intention to perform a certain type of activity or planning for task execution. The alternatives are straightforward. Is forming an intention a separate mental operation? Or are forming an intention and planning for execution part of a singular process? To determine whether this activity could be differentiated, Haynes et al. first had subjects choose which task to perform (add or subtract two numbers) before they were given the numbers themselves and required to select the answer. They then searched in multiple prefrontal areas for a distinctive pattern during the delay corresponding to the subjects' choice to add versus subtract. They identified an area in the medial prefrontal cortex within which distributed patterns for intending to add or subtract could be differentiated. From these, they could predict the subjects' later response with 71% accuracy. Moreover, these distinctive patterns were no longer present during task execution, although an area more posterior was active in execution.

From this evidence, Haynes et al. concluded that intention was encoded separately from preparation of action: "Our new findings resolve this crucial question by showing for the first time that prefrontal cortex encodes information that is specific to a task currently being prepared by a subject, as would be required for regions encoding a subject's intentions" (p. 324). They also found activity in lateral prefrontal cortex from which they could make an above-chance prediction of subsequent action, and so their ultimate claim is that a network of areas is involved in task-specific representations. They suggested further that the contributions of medial and lateral areas might be differentiated, with the medial region representing the subject's choosing the action, since other studies had indicated a role for medial prefrontal cortex in making choices.

So clearly there are some specialized regions, but they are nothing like modules (otherwise a 71% prediction would be too low).

Loosemore and Harley complain that

... this study, like many of the others, gives us information that seems to be locked in at the neural level alone, without coming up to the functional level and telling us something about how the mechanism of “intending to do an action” actually works. Both empirical conclusions—about the distributed spatial patterns, and the change of location between intention and execution—are just giving us different kinds of location data without saying what kind of mechanism is operating, and how it is doing its work.

Loosemore and Harley seem to be looking for a mechanism at a lower level than Haynes et al. were addressing—asking how intending is accomplished, whereas the goal was simply to demonstrate that intending could be distinguished from specific motor planning. This is the apparent conflict between the idea that fMRI reveals modules and the idea that there is substantial interaction (commonplaces 1 and 3). Given this simple goal, there would be no reason to think that the study could have revealed the “mechanism” of intending. It could have turned out, though, that there was no difference in the areas that encoded a specific intention during the delay interval and during execution. Although a researcher might have still argued, in that situation, that the two operations were different but performed in the same brain area, that seems less plausible given the ability to read the intention off the pattern of activation. The more plausible conclusion would have been to reject the distinction between intention and planning the actual response. Further evidence of their interest in functional organization at this level is Haynes et al.’s attempts to link the medial and lateral activity with other results suggesting choosing to specifically involve medial prefrontal activity.

Knoch, Pascual-Leone, Meyer, Treyer, and Fehr

Although Loosemore and Harley claim their paper focuses on brain imaging studies, two of the studies they describe involve other techniques. We do not object to the inclusion of other studies, since we believe multiple means of accessing complex functions is desirable. Knoch, Pascual-Leone, Meyer, Treyer, and Fehr (2006) employ transcranial magnetic stimulation (TMS), which temporarily suppresses the operation of a brain area. The main reason they use a technique to suppress the operation of an area that had been shown in neuroimaging studies to be active when subjects are performing a task is to analyze functionality (commonplace 6). Inhibiting a region can be revelatory (cf. Bechtel and Richardson 1993; Craver 2007). Whereas fMRI imaging can reveal whether blood flow is altered when subjects perform a task, it leaves open the possibility that the neural processing was directly involved in the task (as opposed to being a downstream effect of a region more functionally involved). If task performance

is altered in a manner indicative of a missing operation when the area is suppressed, one has more compelling evidence about the operation the area performed.

The cognitive activity Knoch et al. explored shows up in what is called the “ultimatum game.” Humans often reject deals they find to be unfair even if they thereby end up worse off. (Economists and decision theorists claim this to be an irrational pattern in human reasoning.) This tendency of humans (not chimpanzees) to reject unfair deals even at personal cost is well established in the behavioral literature, and imaging studies relatively consistently show activity in the dorsolateral prefrontal cortex (DLPFC) as well as the anterior insula when subjects are deciding whether to accept the deal (Sanfey et al. 2003). To determine whether the DLPFC was critically involved in evaluating fairness, Knoch et al. suppressed it with repetitive TMS (rTMS). They also introduced a further manipulation—in some trials the unfair deal was chosen by the computer, in some by another human being who stood to profit. Presumably, subjects do not regard an “offer” by a computer as fair versus unfair, though they do assess offers by human subjects in these terms.

Knoch et al. found that impairment of the right DLPFC resulted in higher rates of acceptance of unfair deals proposed by other humans than when it was not suppressed (even though subjects still judged the deals to be unfair). It also eliminated the increased reaction time subjects showed when confronting unfair deals. But it had no effect on the rate of acceptance of unfair deals selected by the computer, where judgments of fairness do not come into play. These results not only demonstrated that the right DLPFC was contributing to the rejection of unfair offers, it also enabled the researchers to reject the alternative hypothesis Sanfey et al. had offered as to the operation it performed—controlling the emotional impulse (originating elsewhere) to reject the unfair offer. If that were correct, under rTMS, subjects should have been even more inclined to reject, rather than to accept, unfair deals. They interpreted the results they did obtain as indicating that the right DLPFC served to override selfish impulses in the service of culturally based conceptions of fairness (for which Henrich et al. [2001] had argued on anthropological grounds).

Loosemore and Harley view this study as at level 2, concerned only with the absolute location of a function. This misrepresents the goal of the study—to establish the inhibitory operation performed by right DLPFC and to decide between two processing models in which inhibition is viewed as operating on two different processes. In applying their own model to these results, Loosemore and Harley seem to simply accept Knoch et al.’s conclusion that DLPFC plays a role in inhibiting selfish responses (that we do not doubt), though not recognizing that this was the main question the study addressed. Loosemore and Harley also mischaracterize Knoch et al., construing them as treating the DLPFC as a “gate” (a term that never appears in the Knoch et al. paper) “specialized to do the job of enforcing fairness.” Knoch et al., however, explicitly construe the DLPFC as “part of a network that modulates the relative impact of fairness

motives and self-interest goals on decision-making.” Fairly clearly, they are thinking in terms of a network of brain areas and seeking only to determine what operation one known brain area performs (commonplaces 3 and 7).

In fact, the alternative accounting Loosemore and Harley offer does not seem all that different from the one Knoch et al. put forward—they suggest that the DLPFC houses “atoms involved in difficult decisions,” and when these are inhibited, the default simple, selfish action is pursued. They say little about what it would be to be involved in difficult decisions—it simply mediates between the “abstract knowledge of unfairness” and the pursuit of the default action. That is, it is involved in applying the abstract knowledge to inhibit the default selfish action, exactly what Knoch et al. set out to demonstrate in countering the interpretation of Sanfrey et al.

Quiroga, Reddy, Kreiman, Koch and Fried

The study by Quiroga, Reddy, Kreiman, Koch, and Fried (2005) is also not an imaging study. It has as a backdrop other work that shows selective responses in the human medial temporal lobe to faces and objects (e.g., Kreiman et al. 2000). In the current study, electrodes were implanted in eight subjects within the amygdala, the hippocampus, the entorhinal cortex, and the parahippocampal gyrus. (The purpose was diagnostic, intended to identify foci for epileptic seizures.) They previously had seen that some of these implants recorded a consistent pattern of response, so that, for example, one showed a response to Bill Clinton’s face from several different perspectives. The researchers suggested that these “neurons might encode an abstract representation of an individual” (Quiroga et al. 2005, 1102). One alternative hypothesis they considered was that the representations might be distributed, depending on patterns of simple features. Thus, they seem to have two hypotheses in mind: (1) that neuronal responses are to fairly specific “low-level” features and the responses to faces are distributed responses; and (2) that neuronal responses are “abstract.” The fact that responses were relatively independent of the perspective does not fit well with the former view.⁵

Loosemore and Harley suggest this is a level 2 or level 3 study, presumably because of the specificity of response Quiroga et al. observe. Loosemore and Harley tell us that the “sparse encoding” endorsed by Quiroga et al. requires a very specific response, so that “a neuron that fires strongly in response to Jennifer Aniston’s face cannot also respond to the faces of the (superficially similar) Julia Roberts or (thematically related and quite similar) Courtney Cox” (pp. 14–15). However, Quiroga et al. are careful to observe: “We do not mean to imply the existence of single neurons coding uniquely for discrete percepts” (p. 1106) (commonplace 4). Some units actually responded to pictures of different individuals; and even if a probe records a response to only one face in the test set, that does not show it would not also respond to other faces.

The actual point is that neuronal encoding/response has the interesting feature that, though the response is to a person, it is “abstract” insofar as it is a response to the face

from many perspectives, including drawings, and even the person's name itself.⁶ Of course, as we've pointed out, it's one thing to show that there's a response to, say, Jennifer Aniston from various perspectives; it's another to show that no other stimuli could elicit a response (commonplace 4). In any case, the former is what is important to make the researcher's fundamental point, which concerns the "invariant" response of some units to some visual stimuli independent of significant variation in the visual features. Again, the purpose is to distinguish the functional response of the neurons. The point is surely not simply to identify the location of some unit, but to discriminate the possible functional analyses.

The key methodological complaint from Loosemore and Harley about this study is how unlikely the result is. It would, Loosemore and Harley suggest, be unlikely to find so many hits. Quiroga et al. say that some images were chosen after an interview with the subjects. Presumably, these images were chosen because of some interest the subject had. If you found that the subject was a Johnny Cash fan, you wouldn't pick pictures of Jennifer Aniston. Loosemore and Harley spend a good deal of time pointing out how unlikely it is to find the right pictures, but that would make sense only if, for example, the images were a random draw from the Internet. That's not at all what the paper suggests.

In their alternative molecular account, the active atoms will tend to be instantiated in the same areas; and if this is so, then the same atom would tend to be activated in the same place in response to the various pictures of Jennifer Aniston. Whatever the merits of their molecular account, this seems simply to embrace the conclusion of Quiroga et al., that the representations are relatively abstract.

General Observation Regarding Imaging Studies

Our defense of construing these studies as contributing to the functional decomposition of mental processes is not meant to deny that it is possible to challenge the functional interpretation of any of them. We do not know, or pretend to know, whether any of these studies will stand the test of time. Neither do we intend to offer a blanket defense for the use of fMRI, or any other imaging techniques. We actually think that convergence of multiple independent lines of evidence is likely to produce results in cognitive science (Henson 2006; Poldrack 2006). Functional MRI is no magic bullet. It is a tool in an arsenal. The methodological moral is perhaps crucial to our takes on the specific studies. We see them as focused on resolving specific issues, and not on determining where cognitive functions are "realized." The pertinent challenges for researchers are those that advance well-defined and conceptually motivated competing accounts of the operations involved in generating the cognitive phenomena. Indeed, most of these studies were conducted to evaluate just such competing accounts. Progress in developing mechanistic explanations often stems from discovering that existing proposals for mechanisms are overly simplified and require incorporation of additional

parts, operations, and more complex organization. (In Bechtel and Richardson [1993], we endeavored to show through examples how such progress is often obtained.) The studies Loosemore and Harley criticize, as well as much work using neuroimaging, is engaged in just this quest.

Conclusion: The Use and Abuse of fMRI

We offer no grand conclusions. We intend no general endorsement of fMRI studies, though we do embrace the technology as one among many. We also recognize that as the power of fMRI increases, and the techniques for using fMRI improve, it will become increasingly useful. We do offer a cautionary moral, based on the sort of commonplaces we began with: fMRI studies are often overinterpreted, by both critics and enthusiasts. We've focused here on one pair of critics, Loosemore and Harley, but we see similar threats from enthusiasts. Although we think Loosemore and Harley offer challenges that are important in locating overinterpretations of fMRI research, we do not think they manage to address the specific concerns those studies aim to resolve. This should not be taken as an endorsement of the specific outcomes of the focal studies, but an acknowledgment of their focus and limitations. From a philosophical perspective, the problem is that the specific studies are targeted at resolving very specific questions, whereas Loosemore and Harley treat them as much more general. These studies are pieces of what Kuhn called "normal science," working within a paradigm.

Within that paradigm, there are commonplaces that Loosemore and Harley ignore. Fundamentally, we think the tendency to ignore the commonplaces follows from a mistaken picture of the character of interlevel research programs and the history of science. Within philosophy, this picture treats the history of science as a tendency toward greater articulation, and progress as a matter of reduction. In this picture, we look to the more fundamental sciences to capture the results of higher-level inquiries. In Loosemore and Harley's criticisms of fMRI studies, this is reflected in the thought that one must have a well-articulated and well-confirmed psychological theory before it is worthwhile to conduct research into the neuroscientific details. The point of the neuroscience would then be to reduce, or capture, the psychological results. It should be clear that we think waiting for a completed psychology, or even the outlines of a completed psychology, is likely to be like waiting for Kafka's gatekeeper. In other cases, we think that the idea that fMRI, or other work driven by techniques from the neurosciences alone, is sufficient to reveal cognitive mechanisms suffers from the same sort of errors. We do think there is a better picture of interlevel research programs that captures both the history of science and the conduct of research involving fMRI. We've certainly not enforced that alternative picture here, though we've suggested how it might generate a different take on the character of fMRI research.

Notes

1. The exception is, of course, information derived from traumatic insults (or surgical ablations) that result in sometimes quite dramatic deficits associated with massive damage to the brain. The fact that these did contribute to our understanding of cognitive function should be enough to give us pause concerning the usefulness of neuroscientific information to cognitive science. We devote some space on this in Bechtel and Richardson (1993).
2. In what is called *reverse inference*, the activation of a brain region is used to infer that some cognitive operation is happening. Such inferences are problematic, and depend crucially on how selective the region is (Poldrack 2006); even when they are legitimate, however, this should not be understood to imply that the region “realizes” the cognitive function. See the discussion of Haynes et al. (2007) in the text.
3. Vision once again offers a useful exemplar. Distance is discerned not only by binocular disparity, but by such things as shift in hue. This reveals itself in a variety of interesting illusions, as Richard Gregory has illustrated.
4. Loosemore and Harley offer a metaphorical rendering of the idea of a bottleneck that would make anyone skeptical; but that doesn’t play any role in the Dux et al. study. It is invented by Loosemore and Harley.
5. On the surface, this result appears to be in tension with Barsalou’s approach to concepts discussed earlier—abstract representations are not, on the face of it, modal. Resolving this and related issues is a topic for further research; we would note, though, that posing such tensions to be resolved is an important benefit of adopting multiple research strategies, neural and behavioral.
6. The most striking case, we think, is actually the responses to Halle Berry, which included responses to displays of her name, her face in various poses, a pencil sketch, and her in the Catwoman costume. In terms of features, there is very little common ground.

19 What Is Functional Neuroimaging For?

Max Coltheart

Functional neuroimaging consists of imaging brain activity using positron emission tomography (PET), functional magnetic resonance imaging (fMRI), or MEG while the person whose brain is being imaged is performing some cognitive task. Scrutiny of the functional neuroimaging literature suggests that such work has so far pursued three (not mutually exclusive) goals:

1. **Neuranatomical localization of cognitive processes** Here, the goal is to discover something about the role of particular brain regions in cognitive processing, by seeking “to determine which particular brain regions or systems exhibit altered activity in response to the engagement of particular cognitive, emotional, or sensory processes” (Poldrack, this volume, p. 147). Loosemore and Harley (chapter 17, this volume) use the term “level 2 studies” to refer to this kind of functional neuroimaging work.
2. **Testing theories of cognition** Cognitive-theory testing using neuroimaging data can take either of two forms. Some functional neuroimaging studies are concerned with one particular theory expressed in cognitive terms, and seek to test that theory. In contrast, other neuroimaging studies are concerned with competing theories expressed in cognitive terms, and seek to adjudicate between such theories. Loosemore and Harley (chapter 17, this volume) use the term “level 4 studies” to refer to this kind of functional neuroimaging work.
3. **Testing neural models** A neural model (Horwitz et al. 1999) is a proposal as to which regions of the brain are activated when some task is being performed, what pathways of communication are used between these regions, and what the functional strengths of these pathways are (see Horwitz et al. 1999, figure 1 for an example). If the model is purely neural, it says nothing about what processing function is performed by each of the brain regions, just as, if a model is purely cognitive, it says nothing about what brain region is used to perform each of the processing functions it proposes.

Poldrack (chapter 13, this volume) refers to localization (goal 1 above) as *the* goal of functional neuroimaging, which might be taken to express the view that studies using