

What Neuroimaging and Brain Localization Can Do, Cannot Do, and Should Not Do for Social Psychology

Daniel T. Willingham and Elizabeth W. Dunn
University of Virginia

Interest in bridging social psychology and neuroscience has seen a significant upsurge. Much of this interest has centered on *brain localization*—the attempt to relate psychological events to locations of brain events. Although many articles have sought to localize brain activity that supports social behavior, scant attention has been paid to the specific methods to be used in integrating brain localization data into psychological theory. The authors describe 4 strategies psychologists can use to integrate brain localization data and psychological theory, and they consider whether social psychology presents special considerations in the use of these strategies. They conclude that brain localization offers a useful tool for some but not all problems in social psychology, and they discuss the types of problems for which it may and may not prove useful.

The purpose of this article is to examine why social psychologists should be interested in localizing psychological processes in the brain. The intersection of neuroscience and social psychology has been vibrant for many years, and substantial progress has been made in understanding the relationship of the social mind and the body's physiology (for a number of examples, see Cacioppo, Tassinary & Berntson, 2000). There is, however, a new enthusiasm for the integration of neuroscience and social psychology. This increased interest may be adduced by the recent (2001) request by the National Institute of Mental Health for applications that combine social psychology and neuroscience and by the increasing number of social neuroscience presentations at conferences. This increased interest has been noted in professional journals (in addition to this special section of the *Journal of Personality and Social Psychology*, see Adolphs, 1999; Ochsner & Lieberman, 2001) and in the press (Azar, 2001). Some change in zeitgeist seems to be at hand.

What is behind this exuberance? We think it is clear that the apparent applicability of functional magnetic resonance imaging (fMRI) and its greater availability to researchers is fueling much of the excitement. fMRI is a method of localizing neural activity that is associated with mental activity. Because it is noninvasive, relatively inexpensive, and can be applied to healthy brains, it is indeed a breakthrough technology. Naturally, fMRI and other methods of localization are but one type of neuroscientific data of interest to social psychologists. The nervous system has different

components (e.g., central vs. peripheral), and each component has more than one measurable characteristic. As shown in Table 1, social psychologists have already made good use of several neuroscientific methods.

Space limitations obviously preclude a detailed examination of all of the methods listed in Table 1, and indeed, that is not our purpose. We focus on anatomic localization because we believe that these data are currently receiving the most attention. Previous treatments have more often focused on reviewing past achievements while encouraging continued work (e.g., Adolphs, 1999; Klein & Kihlstrom, 1998; Ochsner & Lieberman, 2001). We consider two points. First, how does one actually *use* brain localization data to advance psychological theory? Second, do social psychological questions pose different problems or opportunities for the use of brain localization data than other psychological questions?

What is the Point of Brain Localization?

What is Brain Localization?

Brain localization refers to the association of psychological functions with particular structures in the brain. Three techniques provide the bulk of localization data: single-cell recording studies, lesion studies, and functional imaging studies.

In a single-cell recording study, the activity (i.e., rate of action potentials) is measured directly from an individual neuron or from a small group of neurons, almost always in a nonhuman animal. Through a simple surgery, a probe is placed at the brain site of interest that allows this recording. After the animal recovers from the surgery, the activity of the neuron or neurons may be recorded as the animal is awake and freely behaving. The experimenter varies the stimuli to which the animal is exposed or, by training, the responses that the animal executes. In so doing, the researcher seeks to determine an association with what stimuli or behaviors the neural response is maximized and thereby infer what the neuron codes.

Daniel T. Willingham and Elizabeth W. Dunn, Department of Psychology, University of Virginia.

Preparation of this article was supported by National Institutes of Health Grants R01 NS40106-01 and NIH R01 MH065598-01, by National Science Foundation (NSF) Grant SBR-9905342, and by an NSF Graduate Research Fellowship to Elizabeth W. Dunn.

Correspondence concerning this article should be addressed to Daniel T. Willingham, Department of Psychology, 102 Gilmer Hall, Box 400400, University of Virginia, Charlottesville, Virginia 22904. E-mail: willingham@virginia.edu

Table 1
Behavioral Neuroscientific Methods

Part of nervous system	Characteristic studied	Methodology
Central	Location of neural activity	Functional magnetic resonance, positron emission tomography, single-cell recording, electroencephalogram, magnetoencephalogram, lesion studies
Central	Timing of neural activity	Electroencephalogram, magnetoencephalogram
Central	Neuromodulatory (i.e., chemical) effects	Microdialysis, microinjection, perfusion
Peripheral nervous system: Autonomic	Behavioral changes resulting from system activity	Pupil response, respiration, heart rate, electrodermal activity
Peripheral nervous system: Cranial nerves	Behavioral changes resulting from system activity	Electromyography

In a lesion study, a patient with damage to the brain (because of disease, surgery, or other loss) is characterized in terms of the abilities that are impaired and those that are spared. The assumption is that the impaired abilities rely on the brain locations that have been lesioned; that is, if a patient with a lesion to dorsolateral frontal cortex shows a deficit on a working-memory task, it is tentatively concluded that dorsolateral frontal cortex contributes to working memory. Naturally, a number of different deficits (e.g., loss of language, loss of attention, etc.) might lead to a deficit on the test, so careful testing to isolate the affected ability and to rule out alternative candidates of affected psychological processes is crucial.

The third technique is functional brain imaging, for example fMRI, positron emission tomography (PET), electroencephalogram (EEG), or magnetoencephalogram (MEG). PET and fMRI provide indirect measures of blood flow, on the assumption that neural activity creates a demand for increased oxygen or glucose at active regions, which is met by an increase of blood shunted to the

region. EEG and MEG measure products of neural activity, either electric or magnetic. Because tasks of even minimal complexity change the neural activity through much of the brain, researchers compare carefully selected tasks in an effort to isolate the activity associated with a particular psychological function. For example, one might compare the activity associated with reading a word to reading a nonword in an effort to isolate activity associated with lexical access.

A host of problems are associated with localization techniques, including the logic they use, their susceptibility to artifacts, and so on; some of these problems are listed in Table 2. These problems T2 have been described in detail elsewhere (e.g., Phelps, 1999; Sarter, Berntson, & Cacioppo, 1996) and are not the focus of this article, so they are not discussed further. Although formidable, they are clearly not insurmountable, because researchers have made substantial progress in localizing psychological functions in the brain. The first question we address is why psychologists would have any interest in that effort.

Table 2
Common Problems of Interpretation in Localization Methods

Method	Problem
Single-cell recording	Observed activity may be correlated with but not contribute to the process of interest.
Single-cell recording	Interpretation of correspondence between activity and task may be complex; observed activity may be a small part of a complex network.
Single-cell recording	Recorded neuron may receive input from neurons that perform the task of interest; recorded neuron may be "downstream" from neurons actually performing the function.
Lesion	Complex interactions may produce deficits that do not resemble intact function.
Lesion	Deficit may be due to loss of fibers passing through lesion site, not damaged neural tissue at the site.
Imaging	Subtraction technique assumes correct task analysis.
Imaging	Subtraction technique assumes pure insertion of component processes.
Imaging	Standardization of images to increase signal-noise ratio can introduce artifacts.
Imaging	One area may serve different functions depending on input or overall pattern of activation.

Why Localize?

On the one hand, it seems obvious that studying the brain should be useful if one is interested in the mind. The phrase one often hears is that knowledge of the brain “provides constraints” on theories of the mind. But how, specifically, do these data provide constraints, and how do you figure out what these constraints are? In preparation for this article we spoke to several brain imaging experts and asked the question, “How do brain imaging data help a psychologist develop or test a psychological theory?” The modal response was a soft chuckle followed by some version of the sentence, “That’s a good question.” It is quite obvious what localization is good for if your goal is brain mapping. But what if your goal is to develop psychological theories?

To put the problem another way, consider what a critic might say: “So the amygdala is active when people experience fear. So what? Why are we better off knowing it is the amygdala and not the caudate?” The answer is that brain localization can be helpful in solving a specific problem inherent in most psychological models. Psychological models usually use hypothetical representations and processes that operate on those representations. By *hypothetical representation*, we mean a symbol for an entity in the real world. For example, a memory representation of George Bush might contain information about his appearance and would likely be linked to semantically related representations (e.g., a representation of his status as President). More complex representations might incorporate a number of representations within them, for example, a racial stereotype composed of expected physical and personality characteristics and expected behaviors. Representations cannot do anything on their own; processes must operate on representations. In the case of stereotypes, the processes might be activation and application (e.g., Devine, 1989).

Direct evidence supporting the existence of processes and representations cannot be gathered; their existence must be inferred on the basis of their success (once they are incorporated into a theory) in accounting for patterns of data. Therein lies the problem. Hypothetical representations are too powerful, in that it is possible to develop a limitless number of models with them, all of which accomplish the set of abilities one seeks to explain (Hunt, 1999). For example, suppose you have in hand 10 observations about stereotypes—how they develop, when they are applied, and so on. Once you allow yourself the luxury of creating hypothetical processes and representations, it is theoretically possible to create an infinite number of models that account for those 10 observations. How, then, can the theorist select among models?

One strategy is to use response-time data. The competing models that account for your 10 observations will use different processes and representations to do so, and that means that they will vary in efficiency. Some social psychologists (and other psychologists) have taken response time to be an indirect measure of processing efficiency and have used efficiency as a metric for differentiating models.

Brain localization data offer a second way of differentiating models. These data have been helpful by the use of four strategies. Although each is rooted in brain localization, each offers a quite different source of information to the psychologist.

Brain Localization Strategies

Localization Strategy 1: The Psychological Reality of Representations

The first localization strategy entails inferring a representation from neural coding observed in single-cell recording studies. As described above, single-cell recording studies ideally yield data indicating that a neuron or small pool of neurons are maximally responsive when a hypothetical stage of processing occurs. A classic example of such work is Hubel and Wiesel’s (1959) description of neurons in primary visual cortex that are maximally responsive to the perception of simple lines and line intersections. These data suggested that lines could be a useful representation in theories of early vision and indeed inspired a number of such theories (e.g., Selfridge & Neisser, 1960). Thus, the important contribution of the localization technique is a rather direct indication of a representation that contributes to a psychological function.

This strategy has not been directed toward the social realm until relatively recently but has shown some success. Eye gaze is known to be an important social cue, providing information about dominance, attraction, intimacy, and social control (Kleinke, 1986). Single-cell recording studies in monkeys have indicated that a population of neurons in the superior temporal sulcus is crucial for the coding of eye gaze in monkeys (Emery, 2000). Brain imaging studies in humans are consistent with these findings (Hoffman & Haxby, 2000).

Other populations of neurons in monkey superior temporal sulcus code facial expressions. These are not neurons devoted to the identification of individual faces. Rather, they code facial gestures in monkeys that carry social information such as fear or threat (Hasselmo, Rolls, & Baylis, 1989).

Note that these representations need not exist a priori; the visual system is easily powerful enough to calculate gaze direction with general-purpose mechanisms. These data indicate, however, that gaze direction is explicitly represented, and therefore models should use such a representation. The same argument can be made for the recognition of facial expressions. In each case, these data suggest a representation to be incorporated into social psychological models.

Localization Strategy 2: The Psychological Reality of Constructs

The first localization strategy concerns the psychological reality of representations. The second strategy concerns the psychological reality of what we call here *constructs*. For our purposes, a construct is a theoretical component that is a level removed from processes and representations. A construct is composed of a process and representation (or more than one of each) working together to serve a particular function for a larger psychological system. Constructs have a life of their own, because a theory might specify constructs without specifying the processes and representations underlying the construct in great detail. Cognitive dissonance provides an example. Cognitive dissonance is a construct making use of representations (of one’s behavior and one’s self-concept) and processes (comparison process of behavior with self-concept, feeling of arousal; Harmon-Jones & Mills, 1999).

Localization studies using humans are rarely directed at the discovery of a process or representation. Functional brain imaging studies and lesion studies of human patients are usually directed to the localization of constructs such as lexical access (Petersen, Fox, Posner, Mintun, & Raichle, 1988) or motor sequence learning (Willingham, Salidis, & Gabrieli, 2002).

Localizing the brain activity associated with a construct appears to be compelling confirmatory evidence for the psychological reality of the construct. For example, *theory of mind* refers to the ability to consider the thoughts, goals, and intentions of others (Premack & Woodruff, 1978). Some researchers have claimed that this construct has been localized in the brain (Baron-Cohen, 1995), which would appear to be strong evidence for its psychological reality.

A moment's reflection may cause doubt about the power of this evidence, however. Is it not true that any mental act will lead to brain activity that is localizable? Why is it interesting to localize a hypothetical social construct when they are all, in principle, localizable? Suppose my theory uses a new construct I call the *gift lie generator*. This process creates a socially appropriate lie when someone gives you a present that you do not like. To test the validity of the gift lie generator, I put people in an fMRI scanner and tell them to imagine that a close friend has just given them an ugly hat and that they should imagine what they will say to that friend. I observe that the activity of 14 brain sites is reliably associated with my task. Should one conclude that these 14 areas form a network supporting the gift lie generator construct, thus lending greater credence to my theory?

It is likely that one would not draw this conclusion.¹ Rather, one would suspect that what I have called the gift lie generator is, in the brain, a conglomeration of processes such as memory, judgment, emotion, language processing, self-perception, and others. There is not a dedicated brain network to generate gift lies; rather, there is a pattern of activation caused by primitive constructs that are recruited when one performs the particular task I asked subjects to perform. By *primitive* we mean constructs that are not further reducible. So the crucial question for this localization strategy becomes this: How does one know whether a pattern of activation should be described as being due to the construct one hoped to localize or whether the activation is better described as a conglomeration of activity from primitives?

A first (and unsatisfying) answer might be that one knows one has localized a psychological primitive when it is supported by a single brain area. If researchers observe five areas supporting the recognition of facial emotion (e.g., Adolphs, 2002), should they take that as evidence that recognition of emotion is not viable as a theoretical construct because it is composed of more basic processes? We believe that doing so would be a mistake, and indeed, researchers have seldom made that argument. A construct could have psychological validity and yet be supported by a network of five areas. Further, the ability to determine that a construct is supported by a single area is limited by the resolution of the brain imaging techniques used. The number of areas in which a construct is localized is not a reliable index of its theoretical usefulness.

How, then, can one differentiate between the case of emotion recognition, which is deemed viable, and lying about a gift, which is deemed trivial? The answer lies in behavioral testing. There are a great deal of behavioral data lending support to the psychological

reality of emotion recognition as a hypothetical construct (e.g., Ekman & Friesen, 1971) and none supporting the psychological reality of the gift lie generator. But if one relies on behavioral data to decide that a construct has psychological reality, what has localization added?

Localization data can add to confidence that the construct is well described and that the construct is indeed fundamental to social processing—that is, that the construct is not a trivial conglomeration of more basic processes. The data allowing such confidence would be consistent involvement of the relevant brain areas supporting the construct across a wide range of tasks and absence of involvement when the construct is predicted not to be involved in a task. In other words, the amygdala, basal ganglia, and occipito-temporal, orbitofrontal, and right parietal cortices should be involved in any task that requires emotion recognition, and that particular configuration of areas should not be involved in any task that does not. Further, it should be possible to devise a wide range of tasks for both categories.

If the critical brain areas participate when predicted (and do not when not), that finding confirms two things. First, it confirms that the relationship of emotion recognition to tasks is well described. Although the tasks may differ widely, the researcher can predict when emotion recognition will be tapped and when it will not. Second, the fact that emotion recognition is observed contributing to different tasks supports the idea that it is fundamental to social processing. The gift lie generator is a trivial construct because it is not generalizable. It is specific to one task.

This localization strategy has been used to good effect in the study of emotion. Substantial behavioral evidence supported the usefulness of constructs to represent emotions such as anger and fear (e.g., Averill, 1983). Studies using neuroimaging (e.g., Young, Newcombe, de Haan, Small, & Hay, 1993), patient populations (Adolphs, Tranel, Damasio, & Damasio, 1994), and animal models (LeDoux, Cicchetti, Xagoraris, & Romanski, 1990) have localized the anatomic bases of some emotions, and as described as a desideratum above, these studies have shown that the network dedicated to emotion processing operates predictably and reliably across a range of tasks.

Localization Strategy 3: Psychological Separability and Anatomic Separability

Anatomic localization has been used to argue that two apparently similar processes of the mind are in fact separate. The argument is that if X and Y are located in different parts of the brain, then X and Y are psychologically separate. For example, Davidson (1995) has argued that left anterior cortex supports approach-related processes, including positive evaluation of aspects of the environment, whereas right anterior cortex supports withdrawal-related processes, including negative evaluation. This observation indicates the same object or situation can be evaluated

¹ Another obvious problem is that the process of interest might be conflated with some other variable—for example, subjects feel frustrated when thinking of the situation described, so the activation is due to frustration. We are assuming for the moment that careful experimentation and the proper use of control conditions allows the conclusion that the activity is really associated with generating a lie as described.

as simultaneously positive and negative. The alternative would be a single system that makes an evaluation along a single dimension of positivity–negativity, meaning that positive and negative evaluations would necessarily be inversely related.

Fn2

Using this method entails the assumption that anatomic separability implies psychological separability.² The fact that spatial and verbal working memory are supported by different neural structures does not compel different theories at a psychological level of description; it is possible that both types of memory use the same processes and representations even though they are subserved by different brain regions (Weldon, 1999). Nevertheless, most researchers are willing to assume that anatomic separability implies psychological separability.

An important limitation of this strategy is that the relationship between a psychological construct and the brain is not always straightforward. For example, Kelly et al. (2002) asked subjects to make three judgments: whether an adjective described them, whether an adjective described a well-known person (e.g., President Bush), and a control task. The researchers reported that self-referential processing dissociated from other semantic processing; that is, when subjects retrieved information about themselves, this activated different areas of the brain than when they retrieved information from memory about others. Can one conclude that the selective activation is a network for processing information about the self? Such a conclusion would be premature. One problem is that the conditions may differ not only in the crucial to-be-studied factor but also in other ways that are correlated with that factor. For example, subjects might attend more closely to self-referential material. Indeed, the self-referential task might require less attention because the self-referential task entails more familiar comparisons. This problem applies to behavioral studies as well as imaging studies but seems less often remembered in the latter.

A second problem of interpretation applies only to imaging studies. Suppose that knowledge about individuals (including the self) is distributed across different cortical areas, with knowledge of appearance in one place, knowledge of attitudes in another place, and so on. If people know a great deal about their own personality, for example, then that cortical area will be active when they perform a self-referential task because the representation of their own personality is robust but not because there is a dedicated representation of the self anywhere in the brain.

Localization Strategy 4: Brain Informs Theory

In this localization strategy, researchers use existing knowledge of the brain to shape psychological theory. For example, the 1970s saw vigorous debate regarding the fundamental nature of visual imagery (Kosslyn, 1980): Are the representations supporting imagery analog (quasipictorial) or propositional (language based)? If imagery is analog, then it should be supported by brain regions known to support visual perception (e.g., occipital cortex). If imagery is propositional, it should be localized in brain regions known to support linguistic processing (e.g., temporal cortex). Neuroimaging and lesion data localized imagery in secondary and possibly primary visual cortex (e.g., Kosslyn, Alpert, Thompson, & Maljkovic, 1993), strongly supporting the analog nature of imagery.

This localization strategy has been used to good effect in social psychology. For example, Lieberman, Ochsner, Gilbert, and Schacter (2001) tested patients with anterograde amnesia for their response on a standard cognitive dissonance task. The patients showed normal dissonance effects, indicating that explicit memory is not necessary for dissonance processes. Knowledge of brain structure and function was used to gain insight into a social psychological process. Similar work has examined the role of explicit memory on self-concept (Klein, Loftus, & Kihlstrom, 1996) and the role of episodic versus semantic memory on self-concept (Craik et al., 1999).

Naturally, this strategy is only as good as current knowledge of the brain. For example, Hart et al. (2000) and Phelps et al. (2000) conducted fMRI studies of same- and different-race faces and observed amygdala activation associated with out-group faces. It is tempting to tie these results to those showing amygdala contribution to fear conditioning (e.g., LeDoux et al., 1990) and to suggest that fear is a component of stereotyping. In fact, both research teams were appropriately cautious in interpreting the amygdala activation. Amygdala activation has been associated not only with fear conditioning but also with positive emotions such as amusement (Hamann & Mao, 2002) and perception of happiness (Canli, Sivers, Whitfield, Gotlib, & Gabrieli, 2002), among other possible functions. Certainly, there are parts of the brain for which the function is well enough understood that one can interpret an activation or the results of a lesion with some confidence (e.g., occipital cortex, primary motor cortex, primary auditory cortex). There are, of course, many parts of the brain where such confidence would not be warranted.

Characteristics of Social Psychology Bearing on Localization

All in all, it would seem that there is every reason for enthusiasm about the integration of neuroscientific data into social psychology. We have outlined four methods of using brain localization, and it would appear that social psychologists are making some use of all four. Should we not expect continued progress?

Certainly, progress will continue. We also believe, however, that there are limitations to the applicability of neuroscientific data to social psychological problems and that these limitations are traceable to the very nature of social processes themselves. We first describe two features that are characteristic of social psychological constructs and then describe how those two characteristics impact the use of brain localization data.

Complexity

We take it as noncontroversial that there is a hierarchical nature to the mind and that different levels of complexity each have a

² This sort of evidence has been also been used to argue for separate systems. Definitions of systems have varied, but they may be taken to be a collection of constructs, all of which serve a particular function. For example, one might define a system dedicated to attitude change that uses the construct cognitive dissonance, among others. Other more complex assumptions are necessary if one wants to make an argument for separate systems on the basis of localization data (for a discussion, see Willingham & Goedert, 2001).

theoretical integrity of their own. At the bottommost level of the hierarchy are processes and representations. At the next level are first-order constructs, which use processes and representations as building blocks (i.e., they have processes and representations embedded in them). Second-order constructs use first-order constructs as building blocks but might also use some processes and representations, and so on.

We maintain that many functions of interest to cognitive psychologists are first-order constructs (perception, attention, memory, etc.). Many functions of interest to social psychologists are second order (stereotyping, impression formation, conformity, etc.), meaning that they have first-order constructs (often cognitive) embedded in them. The particular instantiations of constructs as first versus second order are not critical (e.g., working memory might be considered second order because it is composed of the phonological loop, visuospatial sketchpad, and central executive). What matters is the fact that some constructs have others embedded in them (the hierarchical nature) and the idea that for the most part social processes are higher on the hierarchy than cognitive processes. This characterization does not always apply by field. There are first-order social functions (e.g., social influences on memory), and there are second-order cognitive processes (e.g., problem solving). Nevertheless, we maintain that many of the functions traditionally of interest to social psychologists have cognitive constructs embedded in them. (This is not the only view; for a discussion, see Ostrom, 1984.)

To provide a concrete example, Gilbert and Hixon (1991) showed that the availability of attentional resources moderates stereotype activation and stereotype application. Participants watched a video of a Caucasian or an Asian American woman displaying cards with word fragments that participants were to complete. Participants were likely to complete the fragments with stereotype-appropriate words when the woman was Asian American, unless attention was diverted with a counting task. A second phase of the experiment showed an effect of attention on the later application of a stereotype. Explicating the workings of attention was not the goal of the experiment; the goal was to show that the construct of attention is necessary to a complete explanation of stereotyping. Thus, stereotyping makes use of the construct "attention." We discuss the implications of this putative complexity of social constructs below.

Temporality

The second characteristic of social processes is temporality. Social interaction is an inherently dynamic system—I change my behavior on the basis of what you do (and my anticipations thereof), and you change your behavior on the basis of what I do. A number of researchers have suggested that this interactive element is a defining element of social behavior (Fiske & Taylor, 1991; Ostrom, 1984; Zajonc, 1980). If social behavior is dynamic, a temporal component is an important part of social models.

Implications of These Characteristics for Brain Localization

Imaging and Time

As noted above, social interaction is, by its nature, dynamic. For that reason, it would seem that a temporal component would be a

desirable feature of models of social processes. Brain localization techniques such as PET and fMRI do not provide this temporal information; they provide a record of activation collapsed across some seconds (as few as 3 or as many as 40, depending on the technique). EEG and MEG record magnetic or electric activity and are capable of tracking very rapid changes in these measures—as rapid as millisecond resolution. Unfortunately, the spatial resolution is poor because the measuring devices outside the skull that do the recording collapse information from a large area of the brain, and in the case of EEG, the electrical signal is diffused by cerebrospinal fluid and the skull. It is feasible to use EEG and fMRI simultaneously to obtain good spatial and temporal resolution (Kruggel, Herrmann, Wiggins, & von Cramon, 2001), but this technique is not yet well established.

An example of the difficulty that temporal information presents may be adduced from efforts to use functional imaging to shed light on consolidation. *Consolidation* is the hypothetical process (or set of processes) by which memories become more stable over time. Although there is substantial evidence for consolidation from animal models (e.g., Anagnostaras, Maren, & Fanselow, 1999) as well as evidence from neuropsychological patients (e.g., Kapur, 1999), the temporal nature of consolidation has been impossible to capture using brain imaging. Researchers can image what they think are the results of consolidation (Haist, Bowden Gore, & Mao, 2001) but not the process itself.

Interactions and Task Analysis

All four of the localization strategies discussed above require task analysis. This requirement is most obvious in the case of brain imaging. Because even the simplest tasks change the activity of much of the brain, researchers must use one of two techniques to isolate a putative psychological construct. They can compare two tasks that differ only in that component; for example, one could subtract the activity associated with attending to a crosshair target from the activity associated with attending to a random letter string and argue that the difference between the tasks is that the latter entails visual perception of word forms and the former does not. Another technique is to systematically vary one characteristic of stimuli across trials and seek activity that is correlated with this systematic variation. For example, you could have participants maintain an increasing number of items in working memory and look for activity that is systematically related to the increase in working memory load.

A fundamental problem is that these task analyses rarely account for the possibility of interactions. For example (Friston et al., 1996), suppose in one condition participants are asked to look at a display in which there is motion. In another, participants are shown the same display, but are asked to attend to motion. The experimenter subtracts the activity of the first condition from the second and observes activity in cortical area V5. From a straightforward task analysis, it would seem that the difference between the conditions is attention to motion, so one would conclude that area V5 supports attention to motion. But from other work it is known that V5 does not support attention to motion but an interaction of attention to motion and perception of motion. Thus, neither attention to motion nor motion perception activates V5—only the combination of the two will do so.

This problem—that psychological processes have complex interactions—is now thought of as perhaps more pernicious than was first appreciated, and this appreciation has in fact been spurred by functional neuroimaging results. For example, a rather puzzling result from a number of imaging studies in the 1990s was the vigorous activation during episodic memory and retrieval tasks of frontal areas thought to be associated with working memory, not with episodic memory. Recent views have suggested that these are indeed working-memory processes, but they are coopted into even simple episodic memory tasks (Wagner, 1999). In other words, those doing task analysis of even relatively simple cognitive processes are still faced with complex interactions.

This problem is pervasive in imaging, and it is well understood. Unfortunately, there are a limited number of ways to deal with it (Friston et al., 1996, suggested one). For the purposes of social psychology, the point is this: Localization demands task analysis, and task analysis is complex. That is doubtless due to the fact that a psychological explanation of even simple phenomena such as motion detection will be complex and will include feedback loops, recursion, interactions, and so on. The brain mechanisms supporting simple processes such as motion detection will be complex and will include feedback loops and all the rest. Thus, this is an effort to map a complex psychological system onto a complex biological system. If doing so is difficult with processes like motion detection, how much more difficult will it be when studying processes (e.g., stereotyping) that have other processes (attention) embedded in them?

Can Be Localized Versus Should Be Localized

The foregoing section describes why it is difficult to localize constructs. Still other constructs may simply be unlocalizable because they are emergent properties. An emergent property arises from the interaction of entities, none of which has the property. For example, some researchers believe that consciousness is not localizable to any one place in the brain but rather arises from the interaction of processes distributed throughout the brain, none of which has the property of consciousness (Searle, 1992). Some constructs that are central to social psychology may also be emergent and therefore unlocalizable. For example, the self may be the product of a wide variety of resources such as long-term knowledge, implicit theories, cultural beliefs, and so on. The self may be an emergent property of a widely distributed network and therefore not localizable.

However, suppose that constructs like the self are decomposable. That case brings up a different issue. The fact that a construct can be decomposed does not necessarily mean that it is useful to do so. For example, one might propose that an attitude is composed of memory representations and affect. If one can localize memory and affect in the brain, should one jettison the construct “attitude” from social theories and replace it with memory and affect?

We believe that this question is germane to many classic and contemporary topics of study in social psychology. For example, Latane and Darley’s (1970) model of bystander intervention entails noticing the event, construing it as an emergency, assuming personal responsibility, and knowing how to help. Each step is arguably deconstructable to smaller grained processes; for exam-

ple, assuming personal responsibility is likely to require the use of long-term memory, reasoning, and working memory, at least.

Models that arose in the more recent era of social cognition often have the same property; they use higher level social constructs that are probably not themselves localizable but that could be deconstructed into lower level constructs that might be localizable. For example, Fiske and Neuberg’s (1990) continuum model proposes that person perception begins with initial categorization, followed by confirmation of this initial categorization or an alternative categorization, and proceeds to attribute-by-attribute processing if categorization fails (with each step requiring adequate motivation and cognitive resources). Although this theory specifies the cognitive processes that underlie impression formation and adequately explains extant research findings, its constructs are unlikely to yield to simple neural localization; the process of initial categorization, for example, probably depends on several subsystems, and the specific systems recruited may even depend on the type of categorization made (e.g., whether the person is categorized as a Black man, an assailant, or a store clerk).

So what should social psychologists do? The field is rife with constructs that are unlikely to be easily localizable (e.g., stereotypes, attitudes, attributions, dissonance, group contagion, reciprocity, attachment) but that may well be deconstructed into more basic constructs that are localizable. We believe it would be a mistake for social psychologists to abandon the level of theoretical development that experience has indicated is appropriate for a lower level of theoretical analysis simply because of the charm of brain localizability. Though timeworn, the analogy to the relationship between chemistry and physics is apt. Chemistry uses constructs that are well understood to be decomposable to physical principles. Nevertheless, principles of chemistry are described in the language appropriate to chemistry, not in the language of physics. In the same way, principles of social psychology may be describable in terms of local activity of the brain, but that does not mean that the language of the brain is the appropriate one for theory development (see also Cacioppo & Berntson, 1992). The constructs that social psychologists use have a theoretical integrity of their own and should not be abandoned in favor of constructs that may be localizable but that will not be functional in a social theory. Further, social psychology should reserve its right to develop theoretical constructs that may not be localizable.

The Prevalence of Purely Social Constructs in the Brain

A valid question concerns the proportion of constructs in the brain that are likely to turn out to be strictly social versus those that are general purpose, meaning that they can operate equally well on nonsocial problems (e.g., Bechara, 2002). To put it another way, will social psychologists find that once they are dealing with the representations used by the brain, they do not find much that looks all that social? If you have a complete cognitive system but no constructs that are dedicated to solving social problems, could these processes combine to create the social information necessary, just as we speculated above that the visual system could calculate gaze direction?

We can say with confidence that even if the answer is “yes, a purely cognitive system could create the necessary social information,” the fact is that it does not. We already know that the brain

uses representations that seem clearly directed toward social problems, namely gaze direction and social cues from facial expressions.

How many more social representations or constructs will be uncovered? Any answer will of course be speculative, but we believe that one bit of data indicates a paucity of purely social constructs in the brain; it is rare to find selective social deficits after brain injury. When the brain is lesioned it is not uncommon to find deficits that are highly selective cognitively (e.g., language production, language comprehension, visual motion analysis, etc.).

One sees this sort of specificity of social deficits seldom, if at all. For example, patients with early damage to prefrontal cortex may be intact cognitively but are impaired socially (e.g., Anderson, Bechara, Damasio, Tranel, & Damasio, 1999). This social impairment is not very specific, however, and spans a number of social domains, including failure to complete tasks, follow rules, or exhibit empathy; chronic lying; a lack of friends; risky sexual behavior; and labile and often inappropriate affect, among other problems. The same is true of patients with bilateral amygdala damage; the social impairment is not focused (Hayman, Rexer, Pavol, Strite, & Meyers, 1998; Tranel & Hyman, 1990). When brain abnormality affects social interaction, it seems to affect a large expanse of social terrain; one does not observe patients who do not show conformity effects but are in all other ways socially normal. We are not claiming that such a finding is impossible; we are claiming that it is notably infrequent, relative to selective cognitive deficit.

We raise this point cautiously, because although such specific deficits have not been observed, it is also true that not many people have been looking very hard for them. On the other hand, it is also true that one need not look that hard to see the specificity of cognitive deficits when they occur; the specificity is remarkable.

So what does all this mean? It means that we suspect that there are not all that many constructs represented in the brain that are strictly social. There may be a few more yet to be discovered, representing information that is vital to our species and to those species from which we evolved—a construct calculating position in a social hierarchy for example. On the whole, however, we tentatively suggest that most of social behavior occurs at a level of description that is at least one level of complexity removed from the representation in the brain. If so, the constructs that can be localized in the brain are simply not the best level of description for theories of many of the problems of interest to social psychologists.

Evaluation

The first half of this article is an attempt to clarify exactly how brain localization data can be used in service of social psychology. The second half discusses characteristics of social psychological problems that might pose a particular problem for these methods. The foregoing section should not be taken as a pessimistic evaluation of the prospects for social neuroscience. As noted in the first section of this article, neuroscientific data are already making a positive contribution to social psychology. In closing, we offer two thoughts for the future.

Much of the emphasis regarding social neuroscience has been on the extent to which neuroscientific methods can inform social

psychology. It strikes us that there is at least as much to be gained from information flowing in the other direction. That is, researchers who seek to map brain function have a great deal to gain from social psychologists. Some areas of the brain have proven very difficult to map—for example, virtually all of orbitofrontal cortex and medial frontal cortex. These regions are thought to subserve the very highest reaches human thought and to play a leading role in supporting social behavior. Researchers whose goal is to map the putative functions of the brain area by area have much to gain through collaboration with social psychologists; their behavioral expertise is vital to the effort of mapping these brain areas. For example, social psychologists are experienced in studying dissociative states (e.g., Wegner, 2002), which could prove invaluable in understanding the possible contribution of medial frontal cortex to conscious will, as in *alien hand syndrome*, wherein a patient feels that one hand moves as though it has a mind of its own (e.g., Bundick & Spinella, 2000).

The other conclusion that we believe can be drawn from an analysis of brain localization methods is that these techniques will be applicable to some problems in social psychology but not all. In particular, we believe that some of the topics of interest to social psychologists are not amenable to brain localization techniques because of the complexity of the processes; they have embedded in them subprocesses that interact, and such complex processes are difficult to localize. It would be a pity if, in their justifiable enthusiasm for this powerful tool, social psychologists subtly shifted their research programs to problems that are amenable to brain localization or shifted their theoretical language to constructs that are localizable. Although social neuroscience should be pursued with vigor, it should not be pursued in a way that will diminish or fail to exploit the progress made in past decades.

References

- Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, 3, 469–479.
- Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12, 169–177.
- Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994, December 15). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, 372, 669–672.
- Anagnostaras, S. G., Maren, S., & Fanselow, M. S. (1999). Temporally graded retrograde amnesia of contextual fear after hippocampal damage in rats: Within-subjects examination. *Journal of Neuroscience*, 19, 1106–1114.
- Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience*, 2, 1032–1037.
- Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist*, 38, 1145–1160.
- Azar, B. (2001, January). At the frontier of science. *Monitor on Psychology*, 33, 40–43.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Bechara, A. (2002). The neurology of social cognition. *Brain*, 125, 1673–1675.
- Bundick, T., Jr., & Spinella, M. (2000). Subjective experience, involuntary movement and posterior alien hand syndrome. *Journal of Neurology, Neurosurgery and Psychiatry*, 68, 83–85.
- Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contribu-

- tions to the decade of the brain: Doctrine of multilevel analysis. *American Psychologist*, 47, 1019–1028.
- Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). *Handbook of psychophysiology* (2nd ed.). Cambridge, England: Cambridge University Press.
- Canli, T., Sivers, H., Whitfield, S. L., Gotlib, I. H., & Gabrieli, J. D. (2002, June 21). Amygdala response to happy faces as a function of extraversion. *Science*, 296, 2191.
- Craik, F. I. M., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., & Kapur, S. (1999). In search of the self: A positron emission tomography study. *Psychological Science*, 10, 26–34.
- Davidson, R. J. (1995). Cerebral asymmetry, emotion, and affective style. In R. J. Davidson & K. Hugdahl (Eds.), *Brain asymmetry* (pp. 361–387). Cambridge, MA: MIT Press.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124–129.
- Emery, N. J. (2000). The eyes have it: The neuroethology, function, and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, 24, 581–604.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation: From category-based to individuating processes as a function of information, motivation, and attention. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–108). San Diego, CA: Academic Press.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.
- Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *NeuroImage*, 4, 97–104.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60, 509–517.
- Haist, F., Bowden Gore, J., & Mao, H. (2001). Consolidation of human memory over decades revealed by functional magnetic resonance imaging. *Nature Neuroscience*, 4, 1139–1145.
- Hamann, S., & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *NeuroReport*, 13, 15–19.
- Harmon-Jones, E., & Mills, J. (1999). *Cognitive dissonance: Progress on a pivotal theory in social psychology*. Washington, DC: American Psychological Association.
- Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *NeuroReport*, 11, 2351–2355.
- Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioral Brain Research*, 32, 203–218.
- Hayman, L. A., Rexer, J. L., Pavol, M. A., Strite, D., & Meyers, C. A. (1998). Kluver–Bucy syndrome after bilateral selective damage of amygdala and its cortical connections. *Journal of Neuropsychiatry and Clinical Neurosciences*, 10, 354–358.
- Hoffman, E., & Haxby, J. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3, 80–84.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148, 574–591.
- Hunt, E. (1999). What is a theory of thought? In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 3–50). Cambridge, MA: MIT Press.
- Kapur, N. (1999). Syndromes of retrograde amnesia: A conceptual and empirical synthesis. *Psychological Bulletin*, 125, 800–825.
- Kelly, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, 14, 785–794.
- Klein, S. B., & Kihlstrom, J. F. (1998). On bridging the gap between social-personality psychology and neuropsychology. *Personality and Social Psychology Review*, 2, 228–242.
- Klein, S. B., Loftus, J., & Kihlstrom, J. F. (1996). Self-knowledge of an amnesic patient: Toward a neuropsychology of personality and social psychology. *Journal of Experimental Psychology: General*, 125, 250–260.
- Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, 100, 78–100.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Kosslyn, S. M., Alpert, N. M., Thompson, W. L., & Maljkovic, V. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, 5, 263–287.
- Kruggel, F., Herrmann, C. S., Wiggins, C. J., & von Cramon, D. Y. (2001). Hemodynamic and electroencephalographic responses to illusory figures: Recording of the evoked potentials during functional MRI. *NeuroImage*, 14, 1327–1336.
- Latane, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Englewood Cliffs, NJ: Prentice-Hall.
- LeDoux, J. E., Cicchetti, P., Xagoraris, A., & Romanski, L. M. (1990). The lateral amygdaloid nucleus: Sensory interface of the amygdala in fear conditioning. *Journal of Neuroscience*, 10, 1062–1069.
- Lieberman, M. D., Ochsner, K. N., Gilbert, D. T., & Schacter, D. L. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychological Science*, 121, 135–140.
- Ochsner, K., & Lieberman, M. (2001). The emergence of social cognitive neuroscience. *American Psychologist*, 56, 717–734.
- Ostrom, T. M. (1984). The sovereignty of social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 1–38). Hillsdale, NJ: Erlbaum.
- Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988, February 18). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331, 585–589.
- Phelps, E. A. (1999). Brain versus behavioral studies of cognition. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 295–322). Cambridge, MA: MIT Press.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, 12, 729–738.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515–526.
- Sarter, M., Berntson, G. G., & Cacioppo, J. T. (1996). Brain imaging and cognitive neuroscience: Toward strong inference in attributing function to structure. *American Psychologist*, 51, 13–21.
- Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Selfridge, O. G., & Neisser, U. (1960). Pattern recognition by machine. *Scientific American*, 203, 60–68.
- Tranel, D., & Hyman, B. T. (1990). Neuropsychological correlates of bilateral amygdala damage. *Archives of Neurology*, 47, 349–355.
- Wagner, A. D. (1999). Working memory contributions to human learning and remembering. *Neuron*, 22, 19–22.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Weldon, M. S. (1999). The memory chop shop: Issues in the search for

- memory systems. In J. K. Foster & M. Jelicic (Eds.), *Memory: Structure, function, or process?* (pp. 162–204). Oxford, England: Oxford University Press.
- Willingham, D. B., & Goedert, K. (2001). The role of taxonomies in the study of human memory. *Cognitive, Affective, and Behavioral Neuroscience, 1*, 250–265.
- Willingham, D. B., Salidis, J., & Gabrieli, J. D. E. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *Journal of Neurophysiology, 88*, 1451–1460.
- Young, A. W., Newcombe, F., de Haan, E. H. F., Small, M., & Hay, D. C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. *Brain, 116*, 941–959.
- Zajonc, R. B. (1980). Cognition and social cognition: A historical perspective. In L. Festinger (Ed.), *Retrospections on social psychology* (pp. 180–204). New York: Oxford University Press.

Received September 25, 2002

Revision received January 27, 2003

Accepted April 11, 2003 ■

APA PROOFS