The New School Psychology Bulletin
2009, Vol. 6, No. 1

Print ISSN: 1931-793X; Online ISSN: 1931-7948

Review Report

The Neuroscientific Study of the Self: Methodological and Theoretical Challenges

Cynthia J. Najdowski, B.A.
University of Illinois at Chicago

E. Samuel Winer, M.A.
University of Illinois at Chicago

Neuroscientific research methods, such as brain imaging techniques, have increasingly been applied to social cognitive research efforts and, in particular, to the study of the self. In this essay we discuss the ability of such research to shed light on the emergent, dynamic psychological phenomenon of self. Although neuroscientific tools can be useful for gaining general knowledge about associated underlying structures, a careful consideration of the methodological and theoretical issues discussed herein is necessary to avoid simplifying or reifying the self.

Keywords: research methods, self, self-referential processing, social cognitive neuroscience, medial pre-frontal cortex

To study the self, social and personality psychologists have examined themes in individuals' thoughts, emotions, and memories in relation to both the immediate demands of the social environment and past experiences (e.g., Cervone, 2004; Markus & Wurf, 1987). In recent years, neuroscientific research methods, such as brain imaging techniques, have been increasingly applied to social cognitive research efforts to determine which brain structures are involved in social psychological processes, to elucidate the relations between such structures, and to provide insight into how the structures operate. Neuroscientists as well as social and personality psychologists have implemented these innovative tools in studies of the self.

For psychologists whose primary expertise is not in neuroscience—as is the case for the authors—applying neuroscientific methods to the study of the self seems advantageous. Indeed, Harré (2002) claimed that social cognitive psychology is "based on a mythical version of physical science methodology" (p. 170) and criticized early behavioral psychologists' attempts to adapt the language and concepts of the physical sciences to the study of persons and the mind. Are psychologists finally getting it right by using neuroscientific research methods, or do these methods have similar limitations? In this essay, we first review some promising empirical findings on the self produced through neuroscientific research methods. We then discuss methodological and theoretical considerations that place these findings in proper context. Our goal is not to repudiate the findings initially presented, but instead to reiterate important caveats that must be attended to when using neuroscientific research methods to study the self.

Promising Empirical Findings

Several researchers who have applied neuroscientific research methods to the study of the self have concluded that a specific cognitive structure, the medial prefrontal cortex (mPFC), is involved in self-processes. For example, by comparing neural activation when participants engaged in judgments about themselves, another person, or neutral stimuli, Kelley et al. (2002) concluded that the superiority in memory associated with self-referential processing was due to the recruitment of the mPFC, which was not active when processing other types of information. Research has also revealed that the mPFC is associated with self-monitoring tasks, self-reflection, self-representations, and self-memories (e.g., David et al., 2006; Gusnard, Akbudak, Shulman, & Raichle, 2001; Johnson et al., 2002; Macrae, Moran, Heatheron, Banfield, & Kelley, 2004; Mitchell, Banaji, & Macrae, 2005).

Although activity in the mPFC is associated with self-processing, it appears to be only one constituent of a neural network that supports what we know of as the self. For example, even though Kelley et al. (2002) determined that the mPFC was the critical structure for self-processing, they also found that activity in the posterior cingulate correlated with self-processing (see also Johnson et al., 2002). The orbitofrontal cortex is also implicated, with damage in this region correlating with impaired self-monitoring (Beer, 2006). Further, there is evidence that distinct neural mechanisms are used for processing cognitive and affective components of self-reflection (e.g.,

Correspondence concerning this article should be addressed to Cynthia J. Najdowski, University of Illinois at Chicago, Department of Psychology, Behavioral Sciences Building, Mail Code 285, 1007 West Harrison Street, Chicago, Illinois 60607-7137. E-mail: cnajdo2@uic.edu. We thank Daniel Cervone and Ellen Herbener for their assistance in critiquing this manuscript.

Moran, Macrae, Heatherton, Wyland, & Kelley, 2006). Additional empirical support for this conclusion is derived from Sherer, Hart, Whyte, Nick, and Yablon's (2005) study which found that, in patients with severe traumatic brain injury, impaired self-awareness was not related to lesions in any specific region but was significantly associated with having numerous lesions across multiple regions. Other researchers also have concluded that the self cannot be located in any single neural structure (e.g., Feinberg, 2001; Lieberman, Jarcho, & Satpute, 2004; Platek, Keenan, Gallup, & Mohamed, 2004; Schmitz, Rowley, Kawahara, & Johnson, 2006). So, although some convincing evidence suggests that the self is located primarily at the mPFC, further evidence indicates that distributed networks are required for self-processing.

Methodological Limitations: Do the Tools Match the Task?

As promising as neuroscientific techniques seem for studying the self, several methodological and theoretical limitations make using them a somewhat problematic endeavor. First, results gleaned from neuroscientific methods are typically correlational. For example, data are collected using electroencephalograph (EEG) measures which record the electrical activity of the brain, positron emission tomography (PET) scanning which monitors glucose uptake in the brain, and functional magnetic resonance imaging (fMRI) which measures metabolic changes in the brain. These methods allow researchers to observe where and what level of neural activity is experienced during specific cognitive tasks. Can causal relationships be inferred by mapping such covariations? Harré (2002) defines instruments, as used in the physical sciences, as "devices that change their state under the causal influence of some changing property of the environment in a way which varies systematically with changes in the environment" (p. 170). EEG, PET, and fMRI measures provide this one-to-one variation, with changes in the scanning instruments mapping directly onto changes in neural activity. When participants perform cognitive tasks and changes in neural activity are reflected in the instruments, it is suggested that these corresponding activations are caused by the cognitive task at hand. The tools, however, may not match the task. As stated by Uttal (2002), "the real problem is the bridge between the neural response and cognition, not the one between the neural response and the fMRI image" (p. 378).

In fact, these tasks often scantly resemble the subjective and individual phenomenon that we refer to as self. For example, Kelley et al. (2002) asked participants to judge whether random trait adjectives described themselves, a common manipulation in this field of research. This methodology obscures between-person differences in how individuals describe the self. For example, the trait "dependable" may elicit quite different reactions across people (e.g., depending on how dependable one feels he or she is, the domain considered), and such differences may be reflected in variations across individual patterns of neural activity. Indeed, Miller et al. (2000) reported that individual patterns of neural activity were rather distinct from group-level patterns.

Another consideration is that neuroscientific researchers have employed a wide variety of instruments (e.g., EEG, PET, fMRI), participants (e.g., psychiatric, brain-damaged, or diseased patients), and tasks (e.g., judging trait adjectives, reflecting on one's own attributes). Consequently, results in the neuroscientific literature have been inconsistent. For example, Perrin et al. (2005) reported that EEG and PET data collected from the same participants engaged in the same tasks failed to converge on any specific neural mechanism that was recruited in the recognition of one's own name. Although this criticism could be leveled against a great deal of nomothetic research and methodological pluralism is encouraged, the value of using multiple research tactics depends specifically upon their ability to converge on reliable results.

The inconsistencies reported may also be a product of researchers' subjective judgments. That is, even though this field of research is deceptively "scientific," data obtained from PET and fMRI scans are open to interpretation. Data are transformed and preprocessed, and the formulae for doing so involve many subjective judgments, including determination of time and spatial resolution and the abstraction of statistical models from individual observations. Subjectivity decreases the likelihood that results are replicable. For example, in a meta-analysis of 275 PET and fMRI studies, Cabeza and Nyberg (2000) found that specific cognitive processes could be localized at best to quadrants, and frequently only to halves of the cerebral cortex because activated neural mechanisms were widely distributed across studies.

A further criticism of the study of self using neuroscientific techniques is the comparison of self-processing states to various "control" conditions. Neural activity that occurs during a control task is considered baseline activation, and this typically is subtracted from the level of activation associated with a specific cognitive task. This subtractive analysis masks the fact that cognitive processes activate nearly the entire cerebral cortex. As a result, researchers risk discounting the importance of structures that are initiating parallel processes, or ignoring connections between structures and processes. Further, the validity of such subtractive comparisons depends heavily on

the conditions used to estimate baseline activation, which vary across studies. For example, participants in control conditions may be told to relax and think of nothing in particular, their eyes may be opened or closed, or they may be told to fixate on a target. They may be asked to "rest," which Stark and Squire (2001) argue does not provide an optimal baseline for comparisons because periods of rest are associated with significant cognitive activity. In addition, patterns of neural activity may differ as a function of baseline directives. Of importance, subtractive analyses must be interpreted in the context of how baseline activity might relate to the tasks that researchers use to assess the self. It is also of note that studies sometimes do not conduct manipulation checks to actually assess participants' cognitions during these "control" conditions, and those that do are often based on retrospective self-reports. For example, Fransson (2006) asked participants to retrospectively measure their self-relevant thoughts on a visual analogue scale ranging from 0-100. Asking participants to respond to open-ended questions about their thoughts and coding responses for self-relevance would validate experimental manipulations in these kinds of studies.

As well, self-processing may not be as unique as non-self processing. Kelley et al. (2002) found that the pattern of neural activation in the mPFC that occurred during self-referential processing was more similar to the pattern that occurred during baseline fixation trials than to those that arose when participants made judgments about others. This suggests that self-processing is the default mode of brain functioning; that is, self-processing is what people do absent of any other goal-directed cognition (see also D'Argembeau, Comblain, & Van der Linden, 2005; Gusnard et al., 2001; Gusnard & Raichle, 2001; Schilbach et al., 2006; but see Fransson, 2006).

A final methodological flaw in neuroscientific experiments is researchers' inability, as yet, to develop representative research designs. Brunswick (1956) called for psychologists to sample situations that represent individuals' normal day-to-day lives. Yet, when participants are enveloped in imaging equipment, the testing situation may be so artificial that it bars researchers' ability to generalize their findings to real-world social psychological phenomena.

Theoretical Limitations: Conflating the Tools and the Task?

In addition to the methodological issues that arise when attempting to study the self using neuroscientific research techniques, there are also substantial theoretical concerns. First, theoretical difficulties emerge when one extrapolates from neuroscientific methods to say that the self can be found (e.g., Kelley et al., 2002). For example, consider Harré's (2002) distinction between tools and tasks. If we know that people can whistle and then we learn that the lips are active during whistling, the initial response might be "whistling is in the lips." A reasonable rejoinder might be that "whistling cannot be located in any single bodily structure (such as the lips) because other structures besides the lips are active."[1] That would be right, but incomplete. Even if one did identify all the bodily structures activated by whistling, whistling still is not "in" them (individually or collectively). Whistling is an activity (a task) that can only be performed if one has the various requisite bodily structures (the tools). That is, the relation between the bodily structures and whistling is tools to task, not storage unit to item stored. Following, it is not simply that the self cannot be located in any single neural structure because there are clearly more structures involved (as previously discussed). Instead, the self is not "in" any of them, individually or collectively. Rather, interactions among interconnected neural components result in synergistic processing (e.g., Norris, Chen, Zhu, Small, & Cacioppo, 2004) and "novel system properties" (van Duijn & Bem, 2005, p. 708). Yet, the assumption behind techniques such as fMRI and PET may lead to difficulties in recognizing emergent and dynamic properties that characterize the self, because their use implies a one-to-one mapping of psychological phenomena to underlying neural mechanisms.

Another difficulty associated with studying the self involves the use of personal pronouns or first person language to describe the self, which can reify the abstract entity of perceiver into an agent in its own right. This problem is illustrated by Hume (1748):

> For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never catch myself at any time without a perception and never can observe anything other than a perception. (Book 1, section 4, as quoted by Harré, 2002)

We refer to more than one thing when we discuss the self. That is, in natural discourse a distinction is made between "I" and "me." Whereas the use of "I" implies a self that is the seat of perception, "me" refers to a self that is not "perceived" but rather implies beliefs about a set of attributes and expressions that are typically characteristic of that perceiver. In fact, Travis (2006) reported that activity in the frontal cortex differs for individuals engaged in first-person cognitions compared to individuals engaged

[1] We thank Daniel Cervone for suggesting this analogy.

in third-person cognitions. Travis's results support the importance of distinguishing between "I" and "me," but many neuroscientific investigations of self and self-processing have neglected to make this distinction.

Finally, it is also important for researchers to take context into account. As stated by Kagan (2007):

> Because brain processes, and the psychological states they permit, are dynamic phenomena influenced by the context in which the measurement occurs, it is reasonable to doubt that agents possess broadly generalizable cognitive and emotional functions that, like eye color, are always available for display across settings and stages of development. (p. 15)

For example, self-awareness can be affected by the intended audience (Baldwin & Holmes, 1987) and different group attachments make different aspects of the self salient (Hermans & Dimaggio, 2007). The current experience of self may depend on one's mood, affect, or psychological state (e.g., Kuiper & MacDonald, 1982). Markus and Wurf (1987) posit that the self is a "shifting array of accessible self-knowledge" (p. 306) that is activated by the current circumstances (see also, Wheeler, DeMarree, & Petty, 2007). Thus, the interactive effect of situational constraints and individual factors contributes to the adaptive and flexible nature of what we know of as self. This critique is of course valid in respect to studies of the self in general, but it is particularly relevant when assessing the utility of neuroscientific methods which require that participants either be connected to or contained by machines.

Conclusion

Many of the methodological and theoretical considerations outlined in this essay are applicable to psychological research of any kind. These considerations are particularly germane to research employing neuroscientific methods, however, because the scientific appearance of results gleaned from such techniques make them particularly susceptible to reification. Kelley et al. (2002), whose work has served as a springboard for this paper due to its intriguing results, have clearly taken many of the previous considerations into account. Their understated conclusion belies the title of their work, "Finding the Self?" Specifically, they conclude that self-referential processing is unique due to the recruitment of the mPFC without differences in left interior frontal involvement. That is, that self-referential processing is "unique in terms of its functional representation in the human brain" (p. 791). This is quite possibly true, but it does not mean that the self has been located. It is important to consider this finding, as well as any resulting from an inherently correlational framework, in proper context. Revisiting the example of whistling, one would be able to see a pattern of behaviors (pursed lips, air blown) that only occur during the act of whistling. That is correct, but is that where whistling is? Likewise, the major underlying neural structure active during self-processing is the mPFC, but that is not where the self is. Authors in this literature seem implicitly aware of this manner of critique, but it is a very delicate semantic detail that can quickly be blurred, and thus demands intricate elucidation.

Our goal is to remind researchers of the pitfalls associated with the use of neuroscientific techniques in studies of the self. This review should be considered a resource for researchers interested in studying the self using neuroscientific methods, not a warning against conducting such research at all. Indeed, the self has been difficult to identify and assess using objective methods of any kind. Although we have outlined a number of caveats, we believe neuroscientific techniques can be useful for understanding the enigmatic self.

Clearly, it is important for researchers to always begin with the recognition that the self is embodied, and that the experience of self is a result of the interaction among neural systems and between those systems and the immediate environment, and that such an interaction is irreducible. Understanding the underlying neuroscientific phenomena associated with the process of self is an important line of social and personality research, and thus, it is important to take into account the methodological and theoretical concerns associated with its study. Of importance, the limits of interpretability when employing neuroscientific methods must underlie any program of research. As stated by Bandura (2001), "psychological properties cannot violate the neurophysiological properties of the systems that subserve them. However, the psychological principles need to be pursued in their own right" (p. 19).

References

Baldwin, M.W., & Holmes, J.G. (1987). Salient private audiences and awareness of the self. *Journal of Personality and Social Psychology, 52,* 1087-1098.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52,* 1-26.

Beer, J.S. (2006). Orbitofrontal cortex and social regulation. In J. T. Cacioppo, P. S. Visser, & C. L. Pickett (Eds.), *Social neuroscience: People thinking about thinking people* (pp. 41-62). Cambridge, MA: MIT Press.

Brunswick, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley, CA: University of California Press.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12,* 1-47.

Cervone, D. (2004). The architecture of personality. *Psychological Review, 111,* 183-204.

D'Argembeau, A., Comblain, C., & Van der Linden, M. (2005). Affective valence and the self-reference effect: Influence of retrieval conditions. *British Journal of Psychology, 96,* 457-466.

David, N., Bewernick, B.H., Cohen, M.X., Newen, A., Lux, S., Fink, G.R., et al. (2006). Neural representations of self versus other: Visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience, 18,* 898-910.

Feinberg, T.E. (2001). *Altered egos: How the brain creates the self.* New York, NY: Oxford University Press.

Fransson, P. (2006). How default is the default mode of brain function? Further evidence from intrinsic BOLD signal fluctuations. *Neuropsychologia, 44,* 2836-2845.

Gusnard, D.A., Akbudak, E., Shulman, G.L., & Raichle, M.E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 4259-4264.

Gusnard, D.A., & Raichle, M. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience, 2,* 685-694.

Harré, R. (2002). *Cognitive science: A philosophical introduction.* London: Sage.

Hermans, H.J. M., & Dimaggio, G. (2007). Self, identity, and globalization in times of uncertainty: A dialogical analysis. *Review of General Psychology, 11,* 31-61.

Hume, D. (1748). *A treatise of human nature.* Oxford: Clarendon Press.

Johnson, S.C., Baxter, L.C., Wilder, L.S., Pipe, J.G., Heiserman, J.E., & Prigatano, G.P. (2002). Neural correlates of self-reflection. *Brain, 125,* 1808-1814.

Kagan, J. (2007). The power of context. In G. Downey, Y. Shoda, & D. Cervone (Eds.), *Toward a science of the person: A festschrift for Walter Mischel.* New York: Guilford.

Kelley, W.M., Macrae, C.N., Wyland, C.I., Caglar, S., Inati, S., & Heatherton, T.F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience, 14,* 785-794.

Kuiper, N.A., & MacDonald, M.R. (1982). Self and other perception in mild depressives. *Social Cognition, 1,* 223-239.

Lieberman, M.D., Jarcho, J.M., & Satpute, A.B. (2004). Evidence-based and intuition-based self-knowledge: An fMRI study. *Journal of Personality and Social Psychology, 87,* 421-435.

Macrae, C.N., Moran, J.M, Heatheron, T.F., Banfield, J.F., & Kelley, W.M. (2004). Medial prefrontal activity predicts memory for self. *Cerebral Cortex 14,* 647-654.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38,* 299-337.

Miller, M.B., Van Horn, J.D., Wolford, G.L., Handy, T.C., Valsangkar-Smyth, M., Inati, S., et al. (2002). Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience, 14,* 1200-1214.

Mitchell, J.P., Banaji, M.R., & Macrae, C.N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience, 17,* 1306-1315.

Moran, J.M., Macrae, C.N., Heatherton, T.F., Wyland, C.L., & Kelley, W.M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *Journal of Cognitive Neuroscience, 18,* 1586-1594.

Norris, C.J., Chen, E.E., Zhu, D.C., Small, S.L., & Cacioppo, J.T. (2004). The interaction of social and emotional processes in the brain. *Journal of Cognitive Neuroscience, 16,* 1818-1829.

Perrin, F., Maquet, P., Peigneux, P., Ruby, P., Degueldre, C., Balteau, E., et al. (2005). Neural mechanisms involved in the detection of our first name: A combined ERPs and PET study. *Neuropsychologia, 43,* 12-19.

Platek, S.M., Keenan, J.P., Gallup, Jr., G.G., & Mohamed, F.B. (2004). Where am I? The neurological correlates of self and other. *Cognitive Brain Research, 19,* 114-122.

Schilbach, L., Wohlschlaeger, A.M., Kraemer, N.C., Newen, A., Shah, N.J., Fink, G.R., et al. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia, 44,* 718-730.

Schmitz, T.W., Rowley, H.A., Kawahara, T.N., & Johnson, S.C. (2006). Neural correlates of self-evaluative accuracy after traumatic brain injury. *Neuropsychologia, 44,* 762-773.

Sherer, M., Hart, T., Whyte, J., Nick, T.G., & Yablon, S.A. (2005). Neuroanatomic basis of impaired self-awareness after traumatic brain injury: Findings from early computed tomography. *Journal of Head Trauma Rehabilitation, 20,* 287-300.

Stark, C., & Squire, L. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 12760-12765.

Travis, F. (2006). From I to I: Concepts of self on an object-referral/self-referral continuum. In A. P. Prescott (Ed.), *The concept of self in psychology* (pp. 21-43). Hauppauge, NY: Nova Science Publishers.

Uttal, W.R. (2002). Functional brain mapping—What is it good for? Plenty, but not everything! (Reply to Mal-

colm J. Avison). *Brain and Mind, 3,* 375–379.
van Duijn, M., & Bem, S. (2005). On the alleged illusion of conscious will. *Philosophical Psychology, 18,* 699-714.
Wheeler, S. C., DeMarree, K. G., & Petty, R. E. (2007). Understanding the role of the self in prime-to-behavior effects: The active-self account. *Personality and Social Psychology Review, 11,* 234-261.