The American Journal of Bioethics, 7(1): 49–76, 2007

ISSN: 1526-5161 print / 1536-0075 online
DOI: 10.1080/15265160601064223

Open Peer Commentaries

The Neuroscience of a Person Network

Elizabeth A. Phelps, New York University

In their article, Farah and Heberlein (2007) argue that evidence from neuroscience may help us understand why a specific and non-arbitrary definition of personhood is so difficult. They suggest that we are possessed with an autonomous person network in our brains that may create the illusion of personhood in our minds from a range of stimuli, some of which clearly are not persons, such as geometric shapes moving around a computer screen in a manner that implies a social interaction (Heider 1944). If this is the case, they argue, then our attempts to define personhood may be confounded because the concept, as represented by our brains, "does not correspond to any real category of objects in the world" (Farah and Heberlein 2007, 37).

The proposal of a person network in the human brain is ambitious, and Farah and Heberlein (2007) should be commended for their bold approach. To begin to use neuroscience evidence to inform ethical debates, it is necessary to attempt to characterize the neural representation of complex human concepts. Typical cognitive neuroscience research has focused on understanding the neural representation of stimuli and concepts that are much easier to define, such as simple shapes, faces, words, or scenes (Gazzaniga 2004). With the advent of sophisticated techniques for studying the human brain, we can begin to build on this earlier work to understand the representation of concepts that may have a greater societal importance, including social group bias, moral judgment, cognitive or willful control, and personhood.

Simply because we can start to tackle these more complex concepts, however, does not mean we should lessen the standards we use to evaluate the strength of the neuroscience evidence arguing for a specific proposed neural representation and its qualities. Of course, this standard may be a challenge to ethicists interested in neuroscience who may not always have the expertise or familiarity with the literature to critically evaluate neuroscience data, much like neuroscientists may not be familiar with standards in ethics. As the field of neuroethics develops, it is important we take advantage of the opportunities to learn from our colleagues from different disciplines about the issues in their respective fields because only then can we fully understand the topic at hand. With this in mind, I will briefly review what I believe are the strengths and the weaknesses in the "person network" proposed by Farah and Heberlein (2007) given what we know from previous literature in cognitive and social neuroscience.

The proposed person network operates according to two basic features. The first is if that the person network is separate from the systems representing other things; although "other things" may engage the person network, it is designed for the representation of persons. The second is that the person network is autonomous. Certain triggering stimuli will engage the person network, whether or not the stimulus is actually a person. In their schematic of the person network, Farah and Heberlein (2007) highlight four brain structures: the fusiform gyrus, the medial prefrontal cortex (MPFC), the temporal-parietal junction (TPJ), and the amygdala.

The evidence that a specific brain region may be primarily associated with the representation of persons in a relatively automatic manner is strongest for the fusiform gyrus, which is commonly known as the *fusiform face area*. In her groundbreaking book first published over 15 years ago, Farah (1990) outlined many of the principles governing the representation of faces by the fusiform gyrus based primarily on data from patients with specific brain lesions. Since that time, data from functional magnetic resonance imaging (fMRI) in humans, intracranial recordings in presurgical patients, and studies in other species have confirmed that this region, although it may not respond exclusively to faces, largely satisfies the features proposed by Farah and Heberlein (2007) as comprising part of a person network (Kanwisher 2006).

For the other three regions (MPFC, TPJ, amygdala), however, the evidence that they conform to the proposed features of the person network is less strong. As outlined by Farah and Heberlein (2007), the MPFC is emerging in a number of fMRI studies as a region that yields a relatively greater blood oxygenation level-dependent (BOLD) signal when a research subject is reflecting on qualities that are typical of persons in relation to other objects. A difference in the BOLD response between two conditions is evidence that this region may be more involved in one task than another, but the response does not indicate that the other task does

Address correspondence to Elizabeth A. Phelps, Department of Psychology, New York University, 6 Washington Place, Room 863, New York, NY 10003. E-mail: liz.phelps@nyu.edu

not also recruit this region. BOLD is a relative assessment. The MPFC is a relatively poorly understood region of the human brain, and studies examining other behaviors, such as the processing of reward, have routinely shown preferential BOLD responses in this region as well (Knutson and Cooper 2005). These studies may be isolating anatomically adjacent regions, but this theory has not yet been proven. In addition, recent evidence suggests that the precise region of the MPFC-BOLD response is modulated by how much you think that person is like you (Mitchell et al. 2006), suggesting that this region responds to complex factors beyond personhood. Finally, one of the biggest difficulties in understanding exactly how the MPFC contributes to person perception is the lack of evidence from techniques other than fMRI. Unlike the fusiform gyrus, where lesions lead to a clear deficit in face processing, there is no evidence that lesions of the MPFC result in a deficit in reflecting on self and others. Although the evidence suggests it is possible the MPFC plays a role in a proposed person network, we do not yet have a good understanding of exactly how critical, exclusive or autonomous this region is when representing personhood.

Our understanding of the precise function of the TPJ is even more limited than the MPFC, simply because there are relatively fewer studies investigating the role of this region in representing personhood. As outlined by Farah and Heberlein (2007), however, the few fMRI studies that have focused on the TPJ in mentalizing about others suggest it could play some role in a person network. One interesting aspect of these fMRI studies and other studies is that the tasks employed by the investigators do not only result in relative differences in the BOLD signal in the main region of interest, such as the TPJ (Saxe and Powell 2006). Often several other brain regions show similar BOLD differences, but they are not the focus of that particular study and are often not discussed in any detail. For example, most of the articles that examine the role of MPFC or TPJ in personhood also show similar responses in other brain regions that are not typically thought to be specifically involved in representing social stimuli, such as the posterior cingulate. This finding highlights why it is important to have confirming evidence from multiple techniques in trying to understand the precise role of any specific brain region in representing personhood.

The final brain structure that is included as part of the proposed person network is the amygdala. The amygdala is a region that has been studied extensively across a range of species using a number of techniques. In contrast to the proposal of Farah and Heberlein (2007), the amygdala has been demonstrated to be critical in a number of non-social functions, such as learning that a neutral stimulus predicts aversive consequences or enhancing memory and attention for events that elicit an emotional reaction (Phelps and LeDoux 2005). Although the amygdala has been shown to play a limited role in social responding in humans, it is clear that many of its other functions are similar across a range of species, including species that we think might be unlikely to have a representation like personhood, such as rats. Given this finding, it is hard to argue that the amygdala possesses one of the primary features of the person network, which is that it is separate from representing other things. Rather, the amygdala seems to have a more general role in representing emotional and potentially important stimuli, including people.

I believe that the proposed person network, as it is described in neuroscience terms, is still far from being articulated. I fully agree, however, with one critical aspect of the Farah and Heberlein's (2007) proposal. The psychological and neuroscience evidence clearly demonstrates that we have a predisposition to create the illusion of personhood. We do this automatically, and it takes some effort or may be impossible to change this perception, even when we are consciously aware that it is inaccurate. In this commentary, I have presented some of my thoughts and concerns about the neuroscience model proposed, but this presentation should not detract from the larger picture. Even if neuroscientists have not agreed on every aspect of the neural representation of a complex human concept, this lack of concordance does not mean that we cannot take advantage of the present evidence to gain insight into ethical concerns. Farah and Heberlein may not have presented a network that satisfies all of my concerns as a neuroscientist, but they have highlighted an important concept that clearly has a unique brain representation. We are just quibbling about the details.

REFERENCES

Farah, M. J. 1990. *Visual agnosia*. Cambridge, MA: The MIT Press.

Farah, M. J., and A. S. Heberlein. 2007. Personhood and neuroscience: Naturalizing or nihilating? *American Journal of Bioethics (AJOB–Neuroscience)* 7(1): 37–48.

Gazzaniga, M. 2004. *The cognitive neurosciences III*. Cambridge, MA: The MIT Press.

Heider, F. 1944. Social perception and phenomenal causality. *Psychological Review* 51: 358–374.

Kanwisher, N. 2006. What's in a face? *Science* 311: 617–618.

Knutson, B., and J. C. Cooper. 2005. Functional magnetic imaging of reward prediction. *Current Opinion in Neurobiology* 18: 411–417.

Mitchell, J., C.N. McCrae, and M.R. Banaji. 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 18: 655–663.

Phelps, E.A., and J.E. LeDoux. 2005. Neural systems underlying emotion behavior: From animal models to human function. *Neuron* 48: 175–187.

Saxe, R., and L.J. Powell. 2006. It's the thought that counts. *Psychological Science* 17: 692–699.