January 2007

Personhood and neuroscience: Naturalizing or nihilating?

Martha J. Farah
University of Pennsylvania, mfarah@psych.upenn.edu

Andrea S. Heberlein
University of Pennsylvania

Postprint version. Published in *American Journal of Bioethics - Neuroscience*, Volume 7, Issue 1, January 2007, pages 37-48.
Publisher URL: http://dx.doi.org/10.1080/15265160601064199

Personhood and neuroscience: Naturalizing or nihilating?

Abstract

Personhood is a foundational concept in ethics, yet defining criteria have been elusive. In this article we summarize attempts to define personhood in psychological and neurological terms and conclude that none manage to be both specific and non-arbitrary. We propose that this is because the concept does not correspond to any real category of objects in the world. Rather, it is the product of an evolved brain system that develops innately and projects itself automatically and irrepressibly onto the world whenever triggered by stimulus features such as a human-like face, body, or contingent patterns of behavior. We review the evidence for the existence of an autonomous person network in the brain and discuss its implications for the field of ethics and for the implicit morality of everyday behavior.

Comments

Postprint version. Published in *American Journal of Bioethics - Neuroscience*, Volume 7, Issue 1, January 2007, pages 37-48.
Publisher URL: http://dx.doi.org/10.1080/15265160601064199

Farah MJ and Heberlein AS: Personhood and neuroscience: Naturalizing or nihilating? American Journal of Bioethics – Neuroscience, 2007;7:37-48.

Personhood is a foundational concept in ethics, yet defining criteria have been elusive. In this article we summarize attempts to define personhood in psychological and neurological terms and conclude that none manage to be both specific and non-arbitrary. We propose that this is because the concept does not correspond to any real category of objects in the world. Rather, it is the product of an evolved brain system that develops innately and projects itself automatically and irrepressibly onto the world whenever triggered by stimulus features such as a human-like face, body, or contingent patterns of behavior. We review the evidence for the existence of an autonomous person network in the brain and discuss its implications for the field of ethics and for the implicit morality of everyday behavior.

THE PUZZLE OF PERSONHOOD

Many of our most foundational concepts, on which we construct our understanding of the world, lack clear definitions. For example, concepts such as *space*, *time* and *life* may have a clear enough meaning to be useful in everyday circumstances, but efforts to specify their meanings more rigorously have exposed the complexities and contradictions underlying their apparent simplicity.

The same can be said of the concept of a *person*. In everyday life we have no problem deciding which entities to refer to as persons: human beings generally qualify and other things generally do not. Yet the attempt to specify criteria for personhood has occupied philosophers for centuries. The earliest explicit definition of personhood came from the sixth-century philosopher Boethius, who equated a person with "an individual substance of a rational nature" (Singer 1994). Cognitive capacities such as rationality have remained important features of most subsequent accounts of personhood,[1] including the two most influential accounts of personhood, those of John Locke and Emmanuel Kant.

For Locke, there were three essential characteristics of personhood: rationality, self-awareness, and the linkage of this self-awareness by memory across time and space. In his words, a person is "an intelligent being that has reason and reflection, and can consider itself the same thinking being in different times and places" (Locke, 1997, X). Kant's formulation also includes intelligence, but mainly for its role in enabling one to act morally. At the heart of moral action, for Kant, was the ability to distinguish between persons and things and treat them accordingly. Whereas things may be valued because they are desirable or useful, persons have an intrinsic value, in Kant's terms a "dignity." In his words "... every rational being exists as an end in himself and not merely as a means to be arbitrarily used by this or that will...rational beings are called persons inasmuch as their nature already marks them out as ends in themselves" (Kant 1948, X).

A more modern construal of persons, within the framework of cognitive science and emphasizing information representation, was offered by Dennett (1978). He incorporates the earlier notions of intelligence and self-awareness as necessary attributes of persons, and adds some additional psychological capacities: the capacity to view others as having intentional mental states, to use language, and to be "conscious in some special way" not shared by other animals (Dennett 1978, 270). Dennett suggests that the latter three are mutually interdependent, arguing that the ability to represent a thought like "A desires x" or "B believes y" requires language, and without such linguistic capacity there can be no special consciousness of the human variety.

We thank the members of our laboratory group at the Center for Cognitive Neuroscience for many fruitful discussions regarding personhood and the social brain, and an anonymous reviewer for helpful comments on this article. The writing of this article was supported by R21-DA01586, R01-HD043078, R01-DA18913 and a postdoctoral fellowship through T32-NS07413 at the Children's Hospital of Philadelphia.

Address correspondence to Martha J. Farah, Center for Cognitive Neuroscience, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA 19104. E-mail: mfarah@psych.upenn.edu

[1]Judeo-Christian theology offers a different perspective on personhood, emphasizing a person's relationships, including relationships with other persons and with God. According to this view, it is the participation in these relationships that endows an individual with personhood (Brown 2004). Although interpersonal relationships normally require rationality and many of the other psychological capacities to be discussed, this tradition also recognizes the relationships between humans who lack such capacities and persons who care for them.

Dennett also recognizes the limitations of his list of criteria, however, saying "human beings or other entities can only aspire to being approximations of the ideal [person], and there can be no way to set a 'passing grade' that is not arbitrary" (Dennett 1978, 285).

A number of other contemporary writers have attempted to define personhood, but they have been no more successful at specifying the requirements for, in Dennett's words, a "passing grade." For example, Joseph Fletcher (1979) proposes 15 criteria for personhood. He begins with intelligence, and makes an admirably straightforward effort to specify the dividing line between persons and non-persons by referring to intelligence quotient (IQ) scores: "Below IQ 40 individuals might not be persons; below IQ 20 they are definitely not persons." The problem with this criterion is that, while it is explicit and precise, it is also arbitrary. His other 14 "marks of personhood" include traits and capacities similar to the ones already mentioned as well as a few additions and elaborations. They are: self-awareness, self-control, sense of time, sense of futurity, sense of the past, capacity to relate to others, concern for others, communication with other persons, control of existence, curiosity, change and changeability, balance of rationality and feeling, idiosyncrasy, and neocortical function.

A few other contemporary definitions of personhood will be quoted here for the sake of indicating their fundamental similarities, both in the human traits singled out as relevant to personhood and in the difficulty of translating any of these sets of traits into operational criteria for deciding which entities are persons and which not. From Tooley (1972): something is a person "if it possesses the concept of a self as a continuing subject of experiences and other mental states, and believes that it is itself such a continuing entity." From Feinberg (1980,189): "persons are those beings who are conscious, have a concept and awareness of themselves, are capable of experiencing emotions, can reason and acquire understanding, can plan ahead, can act on their plans, and can feel pleasure and pain." From Englehardt (1986, 107): "What distinguishes persons is their capacity to be self-conscious, rational, and concerned with worthiness of blame or praise." From Rorty (1988, 43): "A person is ... (a) capable of being directed by its conception of its own identity and what is important to that identity, and (b) capable of interacting with others, in a common world. A person is that interactive member of a community, reflexively sensitive to the contexts of her activity, a critically reflective inventor of the story of her life."

Personhood as a Foundational Concept in Ethics

The reason for seeking criteria for personhood is that personhood is a foundational concept in many systems of ethics. Persons, and not other things, are generally held responsible for their actions, and can thus deserve credit or blame. For example, if a person works hard and accomplishes something good, we give that person moral credit. The same is not true for a non-person. If a car revs its engine and moves up a steep hill, we may be pleased at the outcome and value the car more highly, but we do not consider it a morally good car or praise its efforts. Similarly, only persons deserve blame. Indeed, even when the injury caused by a person was not intended, but merely the unfortunate consequence of intentional negligence, common law and the model penal code hold the person responsible. In contrast, and perhaps too obvious to merit comment, we do not assign blame to non-persons. For example, if a falling tree branch kills someone, we do not regard the branch or its behavior as morally wrong.

In addition to being moral agents, and hence responsible for their actions, persons are also moral "patients." Injuring or failing to help a person is morally wrong in a way that similar actions toward other kinds of entity are not. Bioethical discussions of rights generally pertain to the rights of persons (e.g., Universal Declaration of Human Rights 1948). The vague but frequently invoked bioethical concept of "dignity" also seems closely related to personhood and has been defined as "the presumption that one is a person whose actions, thoughts and concerns are worthy of intrinsic respect" (Nuffield Council on Bioethics 2002, cited by Macklin 2003). It is the moral patient aspect of personhood, rather than the moral agent aspect, that has been the focus of much theorizing and debate in bioethics.

The four principles of bioethics, autonomy, nonmaleficence, beneficence and justice (Beauchamp and Childress 2001), apply specifically to persons. For example, in their book, *Principles of Biomedical Ethics*, Beauchamp and Childress refer to the first three principles thus: "Morality requires not only that we treat persons autonomously and refrain from harming them, but also that we contribute to their welfare," (2001, 165) and frame the need for the fourth principle thus: "Standards of justice are needed whenever persons are due benefits or burdens because of their particular properties or circumstances" (2001, 226).

Many of the most contentious issues in bioethics arise in cases involving entities regarded as persons by some and non-persons by others. In such cases it is unclear whether to apply the principles of Beauchamp and Childress. Examples come from both ends of the human lifespan and from ethical issues involving nonhuman animals as well.

Discussions of abortion often focus on the question of whether a fetus is a person and similar questions have arisen in relation to embryos in the context of therapeutic cloning for stem cell research. Similarly, a host of issues surrounding the definition of death and treatment of vegetative patients hinge on differences in our views, not of biological death, that is, the loss of vital functions that sustain the body, but of personal death, that is, the loss of personhood. The difficulty of resolving these issues stems from the lack of defining criteria for personhood.

Finally, although some arguments for improved treatment of animals explicitly deny the relevance of personhood for moral decision-making (e.g., Singer 1979), others focus on it. Extending the legal concept of a person to some species of animal, for the sake of improving animal welfare, is more than a legal maneuver. It is also the expression of a new ontology, that is, a new understanding of what persons are. Just as slaves were once regarded as non-persons,

both legally and more generally in terms of people's beliefs and behaviors, so too authors such as Steven Wise (2002) suggest that certain animals are now wrongly classified as non-persons.

In sum, personhood is a foundational concept in ethics, including both pure philosophical ethics and the applied field of bioethics. Nevertheless, defining criteria for personhood have been elusive. The existence of persons in the world seems intuitively obvious but our intuitions are much less clear on what makes an entity a person. The problem is that, once we have moved from questions of the *kinds* of psychological traits that define persons, for which we have clear intuitions, to more specific formulations, our intuitions abandon us. It is not obvious what is the right subset or hierarchy of traits such as intelligence, language or the ability to represent the mental states of others, nor how well or fully an individual must possess any of these potentially graded abilities. We are left setting criteria that feel, in Dennett's words, arbitrary. In effect, personhood is a concept that everyone feels they understand but no one can satisfactorily define.

NATURALIZING PERSONHOOD

An understandable reaction to the elusiveness of personhood as a metaphysical concept is to refocus our efforts at definition on a more empirical plane. Perhaps there is a "natural kind" in the world that corresponds to persons, and by collecting the right kind of data we can discover its necessary and sufficient properties. More specifically, this approach would seek objective and clear-cut biological criteria that correspond reasonably well with most peoples' intuitions about personhood. These criteria could then be substituted for intuition in those cases where intuitions fail to agree.

This project has the character of developing a scientific taxonomy in place of a folk taxonomy. For example, in everyday life we know the difference between plants and animals. The criteria by which nonscientists would define plants would most likely include being green, not moving, and not killing for food. These criteria work for the most cases, but there are exceptions: Some plants are not green, some move, and some capture insects for food. Biology has revealed a more essential difference between plants and animals, namely that only plants photosynthesize. Perhaps biology can get to the bottom of personhood too, by revealing the essential differences between persons and non-persons.

Within biology, the natural field in which to seek the equivalent of photosynthesis for personhood is neuroscience. The human brain is responsible for the abilities identified by Locke and his successors as crucial for personhood: intelligence, rationality, self awareness, cognition about the future, linguistic communication, mental states of all kinds, including mental states about other people's mental states, and all forms of consciousness.

Accordingly, the abortion debate has been cast by some as an issue of when brain function begins in prenatal development. Several different milestones of neural development have been proposed as the beginning of "brain life" and hence person life (see Jones 1989; Moussa and Shannon 1992). Many of these concern the structure or function of the cerebral cortex of the brain, because it is mainly this part of the brain, in contrast to more primitive structures, which gives rise to the relevant psychological capacities such as intelligence and self-consciousness. Examples of the proposed milestones include the initial formation of cerebral cortex (e.g., Haring 1972) and the first detectable cortical electroencephalogram (EEG) reading (e.g., Gertler 1986).

The major difficulty with this approach is that prenatal brain development is a gradual process, and lacks the kinds of punctate, qualitative transition points that would most naturally be associated with the momentous transformation from non-person to person. Furthermore, many of the milestones that have been proposed as marking a transition depend as much on our technologies for studying fetal brain function as on the fetal brain itself. For example, if we were to measure cortical function by a more sensitive measure than EEG we might choose an earlier gestational age. If we were to measure cortical function more selectively than by EEG, that is using a method that distinguishes different types of neural activity, we might find that cortex does not begin to function as a normal human cortex until a later gestational age. As Green (e.g., 2002) has pointed out, the study of prenatal brain development has not revealed any obvious clefts separating young human non-person tissue from young human persons or even from young human persons-to-be.[2]

At the other end of the lifespan, the concept of "brain death" has met with more acceptance than "brain life," and is the basis for contemporary medical and legal definitions of death. However, brain death, meaning loss of clinically detectable function of the whole brain or loss of function of brain stem structures is not relevant to the question of personhood. As many writers on the topic of brain death have observed, whole brain and brain stem definitions of death correspond to death of the biological human as an integrated homeostatic system rather than to the death of the human person per se (e.g., McMahan 1995). Death of the person is generally associated with loss of higher cortical brain functions, which normally instantiate rationality, self-awareness and the other psychological traits discussed in the previous section. Patients with extensive cortical damage but functioning brain stems are sometimes referred to as *cortically brain dead*. Such patients are more commonly described as being in a *persistent vegetative state*, biologically alive but considered by many to be former persons because they appear to lack any mental life (Jones 2004).

If the psychological traits associated with personhood are largely functions of the cerebral cortex, then naturalizing personhood will require understanding the cortical bases of these traits, a task well underway in the field of cognitive

[2]Green (2002) goes on to suggest that, in the absence of a natural dividing line between prenatal persons and non-persons, we must take an active role in deciding where to draw the line. Our main point, in contrast, is simply that there is no natural dividing line.

neuroscience. The relatively general concepts of rationality and intelligence have long been associated with prefrontal cortex, and recent work has decomposed these psychological capacities into more elementary components such as working memory, inhibitory control, and self-monitoring ability, and localized them more specifically in sub-regions of prefrontal cortex (Miller and Cohen 2001). Cognitive neuroscientists have also made progress in understanding the ability to remember the events of one's life (Squire 2004), to communicate with language (Martin 2003) and even to think about the future (Fellows and Farah 2005).

We believe that this empirical, neuroscience-based approach to defining personhood will eventually be successful in translating the psychological criteria discussed earlier into neurological criteria. In so doing, however, it will be equally successful as the psychological approaches, not more successful. A human with normal brain function may be easy to classify as a person, and a decorticate human may be equally easy to classify as a non-person, but which cortical systems in which combinations are critical and how much functionality is required of each of those systems? Certain cortical systems clearly do not matter; for example, an otherwise normal cortically blind human is still a person. Which systems do matter? Is a globally aphasic patient, who cannot understand or produce language, no longer a person? And assuming it were clear which systems matter, how functional must those systems be? Imagine that a previously healthy human loses a neuron at a time from the critical brain areas until no neurons are left. Relevant clinical observations and neural network modeling indicate that the change in psychological capabilities would be gradual and would in general lack the kinds of qualitative transition points that could be used as non-arbitrary places to draw a line between persons and non-persons (O'Reilly and Munakata 2000). Thus, for defining personhood the devil is just as much present in the neurological details as in the psychological ones.

The real contribution of neuroscience to understanding personhood may be in revealing not what persons are, but rather why we have the intuition that there are persons. Perhaps this intuition does not come from our experiences with persons and non-persons in the world, and thus does not reflect the nature of the external world; perhaps it is innate and structures our experience of the world from the outset. Thus, instead of naturalizing the concept of personhood by identifying its essential characteristics in the natural world, neuroscience may show us that personhood is illusory, constructed by our brains and projected onto the world.

PERSONHOOD, BRAIN REPRESENTATION, AND REALITY

It is fairly widely accepted that we perceive and understand the world using our brains, but this view has important consequences for metaphysics and epistemology that may not be as widely appreciated. We can only understand categories of reality and their regularities and interrelationships if our brains are capable of representing these categories. Assuming that our brains were shaped by natural selection, we might expect a fairly good fit between normal human perceptions of the world and the objective physics of the world that is relevant to our survival. That is, there are good reasons to believe that our perceptions of the size, motion, and temperature of objects map onto the human-scale reality in fairly simple, lawful ways.

Although there is room for variation in this mapping, even this variation supports the more general conclusion that we perceive and understand only what our brains represent. In the perception of sound, some of us perceive absolute pitch while most do not; some of us even perceive sounds as having colors or tastes. Both perfect pitch and synesthesia can also be understood in terms of the ways in which the brain encodes and represents mechanical vibrations in a certain range of frequencies. For example, in the brains of synesthetes, a sound activates not only classical auditory areas, but also areas normally activated by visual inputs (Paulesu 1995). More relevant to the present issue, just as differences between the ways in which different human beings experience the world is attributable to differences in brain function, so too the commonalities among our conceptions of the world are determined by common features of our brains.

The brain represents different types and sources of energy in ways that preserve functionally useful information such as location and intensity. In addition, the physical distinctions among light, heat and sound, for example, are mirrored in the brain's representation of the world, with vision, touch, and hearing implemented in anatomically distinct systems. Indeed, this isomorphism between different aspects of physical reality and brain representation continues at finer-grained levels, with for example the wavelength of light represented in different parts of visual cortex from its location, pattern, or motion (see Farah 2000).

If human survival depends not just on negotiating the physical world but also the social world, then we might expect our brains to have evolved some additional representational "vocabulary" beyond the kinds of physical predicates just discussed. And indeed, one of the most exciting developments in cognitive neuroscience is the discovery of brain systems that appear to be specialized for representing information about people. This research will be summarized next, followed by an analysis of its implications for our thinking about persons.

EVIDENCE THAT WE ARE HARDWIRED TO REPRESENT PERSONS

The earliest clue that the organization of our brain representations carves the world into persons and non-persons came from studies of visual perception in brain-damaged patients. A rare disorder known as *prosopagnosia* consists of impaired visual recognition of the human face (see Farah 2004). Prosopagnosia can be a relatively isolated impairment, that is, a prosopagnosic patient may fail to recognize faces but succeed in recognizing other equally challenging types of objects, consistent with the existence of a

specialized face recognition system that can be damaged selectively (Farah et al. 1995). The recognition of even animal faces may be spared in prosopagnosia, implying that the face recognition system is specialized for representing humans (McNeil and Warrington 1993). The opposite pattern of visual recognition impairment has also been observed, namely generally poor object recognition with preserved face recognition, further strengthening the case for a distinct face recognition system (Feinberg et al. 1994).

Functional neuroimaging of healthy individuals has confirmed the existence of a brain region specialized for human face recognition and localized it with greater precision than is possible with naturally occurring brain lesions (Kanwisher et al. 1997). The fusiform gyrus, on the ventral surface in the brain, is activated disproportionately by the sight of a human face, relative to many other types of visual stimulus materials. Although some controversy exists regarding whether this area is best described as responding to faces *per se* or to a set of perceptual and cognitive demands that are normally associated with face recognition (see Tarr and Gauthier 2000), no one would deny that this area is normally recruited for human face recognition. Figure 1 shows the location of the fusiform gyrus in the human brain. Facial expressions of emotion, as well as vocally expressed emotion, activate additional brain areas including the amygdala (Phillips et al. 2003), also shown in Figure 1. Patients with bilateral amygdala damage are impaired in the perception of people's emotional states (e.g., Adolphs et al. 2005).

Other perceptible aspects of people are also represented by distinct brain systems. Downing and his colleagues have shown that the sight of human bodies, with faces obscured, activates two distinct regions within the brain, one on the fusiform gyrus adjacent to, but distinct from, the face area (Peelen and Downing 2005) and one on the lateral surface of the brain near the temporoparietal juncture (Downing et al. 2001), also shown in Figure 1. Silhouettes and even stick figures of people activate these regions, but equally complex shapes that are not bodies do not.

Bodily movements activate another part of the temporoparietal junction, somewhat anterior to the body area (e.g., Grossman et al. 2000). Studies with "point light walker" stimuli have shown that this region is specialized for the representations of actions *per se* rather than the body; these stimuli, generated by filming in darkness actors who have light emitting diodes attached to various points on their bodies, convey the characteristic motion of a human body while excluding its other visual characteristics (Allison et al. 2000). Parts of the temporoparietal junction are activated specifically by actions perceived to be goal directed (Saxe et al. 2004), and other parts are activated when we think about people's mental states, even in the absence of visual input (Saxe and Wexler 2005).

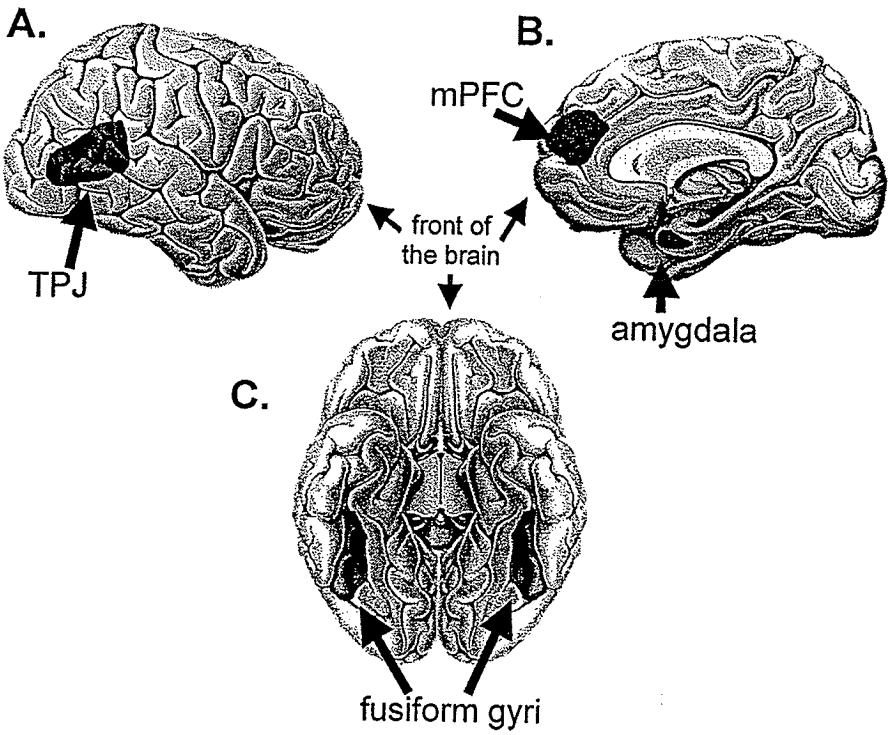

Figure 1. Three views of the human brain: A) a lateral (side) view of the right hemisphere, showing the temporoparietal junction (TPJ); B) a medial (middle surface, between the two hemispheres) view of the right hemisphere showing the medial prefrontal cortex (mPFC) and the amygdala (which is buried inside the cortex but is here shown 'glowing' through); and C) a ventral (bottom) view showing the fusiform gyrus.

Thinking about the mental traits, states and interactions of others also activates the medial prefrontal cortex, shown in Figure 1. In a pioneering study, Fletcher et al. (1995) compared the brain activity evoked by understanding two kinds of stories: stories for which it was necessary to represent someone's mental state to understand the story, and stories for which physical causation rather than psychology had to be represented. For example:

> A burglar who has just robbed a shop is making his getaway. As he is running home, a policeman on his beat sees him drop his glove. He doesn't know the man is a burglar; he just wants to tell him he dropped his glove. But when the policeman shouts out to the burglar "Hey, you! Stop!", the burglar turns around, sees the policeman and gives himself up. He puts his hands up and admits that he did the break-in at the local shop.
> A burglar is about to break into a jeweler's shop. He skillfully picks the lock on the shop door. Carefully he crawls under the electronic detector beam. If he breaks this beam it will set off the alarm. Quietly he opens the door of the storeroom and sees the gems glittering. As he reaches out, however, he steps on something soft. He hears a screech and something small and furry runs out past him, towards the shop door. Immediately the alarm sounds.

The brain activity associated with understanding the two types of story differed in the medial prefrontal cortex. A later study by Gallagher et al. (2000) replicated this localization with similar stories and also with nonverbal cartoons designed to vary in the degree to which they require the viewer to represent the psychology of others.

The medial prefrontal region has been found to represent other aspects of mental processes in a variety of very different task contexts. For example, Goel et al. (1995) compared brain activity while participants judged whether Christopher Columbus would know how to use various objects such as a compact disc to brain activity during other kinds of judgments about the objects, and found greater medial prefrontal activity when Columbus' knowledge was being considered. In a different type of task, Mitchell et al. (2002) asked participants to decide whether a given adjective could ever be true of a given noun. In some cases the adjectives were psychological (e.g., assertive, energetic, fickle, nervous) and could only apply to people, and in other cases they were appropriate for fruits (e.g., sun-dried, seedless) or clothing (e.g., patched, threadbare). Accordingly, nouns were the first names of people, fruits and articles of clothing. Patterns of brain activity associated with judgments about people and non-people were distinct, with a high degree of agreement between the areas associated with person processing in this study and in the previous ones, despite the very different type of task used.

More recently, Mitchell et al. (2005) presented participants with photographs of people and objects, and accompanied each photograph with a statement designed to create a positive or negative impression. For example, a picture of a person might be accompanied by the statement "promised not to smoke in his apartment since his roommate was trying to quit" and a picture of a car might be accompanied by "recently had new fog lights installed." In one condition, participants were told to form an impression of the people and objects based on the statements, and in another condition they were told to remember the sequence in which the statements were presented. The authors confirmed their prediction that impression formation instructions with face photographs would be associated with the most medial prefrontal activation, as these trials involved the most cognition about other persons.

A final example of evidence for dedicated brain systems for representing people comes from a game of "rock/paper/scissors" played in the scanner with a computer whose responses were randomly generated (Gallagher et al. 2002). Participants believed that the responses came from a human in one condition and from a computer in the other. When the conditions were compared, the medial prefrontal cortex was again found to be more active in the human condition.

The weight of the evidence, from a sizable literature only sampled here, clearly supports the conclusion that the human brain represents the appearance, actions, and thoughts of people in a distinct set of regions, different from those used to represent the appearance, movements and properties of other entities. These regions together form a network that is sometimes referred to as "the social brain" (e.g., Brothers 1990; Adolphs 2003; Skuse et al. 2003) but could equally well be termed a network for person representation.

THE AUTONOMY OF THE PERSON NETWORK

In addition to supporting the existence of a separate system for representing persons, recent neuroscience evidence suggests a surprising level of automaticity of person processing by this network, as well as a high degree of innateness. By the term *automaticity* we mean the tendency of the person network to be triggered by certain stimulus features even when we are aware that the stimulus is not a person. By the term *innateness* we mean the genetically preprogrammed nature of the system, without a need to learn that persons exist in the world. The autonomous development and functioning of the person network has important implications for how we think about persons. (Indeed, in some cases this irrepressible autonomy has implications for how we think about non-persons too, as when we coax or curse at our computers.)

Early evidence for the automaticity of face recognition came from a prosopagnosic male patient whose ability to process faces was actually improved when the faces were turned upside-down (Farah et al. 1995). If the faces were shown to the patient in a normal orientation, his damaged face recognition system interfered with his ability to perceive the faces; he was unable to "turn off" his face recognition and treat the faces like some other kind of object or pattern, even though treating them as faces was counterproductive.

Another manifestation of the automaticity of the person network is the ability of certain "trigger features" to engage it. Not just realistic depictions of people but also smiley faces

and stick figures activate the system (Downing et al. 2001; Wright et al. 2002). In other words, we need not believe that a person is present to engage our person network. This is presumably the explanation of a recent finding in behavioral economics, that people adopt more generous strategies in a computer-run economic game when the computer screen happens to display a pair of cartoon eyes (Haley and Fessler 2005). Indeed, the person network could be described as having a hair trigger for these visual features. Faces and bodies can activate the system even when we are not paying attention to them and even when we are unaware of them (Downing et al. 2004; Vuilleumier 2000).

In addition to visual shape features such as static eyes, faces and bodies, certain patterns of motion are also effective at engaging the system. In particular, contingent "behavior," by which a stimulus seems responsive to its environment, can evoke a sense of intentionality and personhood. In the famous animated film of Heider and Simmel (1944) two triangles and a circle move around the screen with motions that are interrelated, giving an impression of three entities interacting with motivations and intentions (see http://pantheon.yale.edu/~bs265/demos/causality.html). The automaticity of this attribution is apparent in the difficulty of describing this film without using psychological terms such as "wants" and "tries" (Scholl and Tremoulet 2000). This automaticity seems related to the triggering of the person network, in that a patient with complete bilateral amygdala degeneration described the film in purely physical terms (Heberlein and Adolphs 2004).

Brain imaging studies of Heider and Simmel-type animations show that all of the brain regions shown in Figure 1 are activated (Castelli et al. 2000; Martin and Weisberg 2003; Schultz 2003). For example, in the study of Martin and Weisberg (2003), two sets of animations were presented: both were composed of moving squares, triangles and circles, which moved in a contingent interactive manner (e.g., as if dancing together or chasing each other) in the "social" set and in a manner consistent with mechanical motions (e.g., like billiard balls or objects on a conveyer belt) in the "mechanical" set. Despite the absence of anything resembling a human being in these animations, the former set and only that set activated the fusiform face area, amygdala, temporoparietal junction and medial prefrontal cortex.

The evidence just reviewed indicates that the person network functions largely autonomously, independent of our conscious, rational beliefs about the nature of smiley faces or animated geometric shapes. Other evidence indicates that its development is also autonomous, in the sense that its specialization for persons comes about prior to experience with persons and other objects in the world.

Evidence for the innateness of the person–non-person distinction comes from the behavior of newborn infants. Johnson et al. (1991) showed that newborns tested within 30 minutes of birth show a greater tendency to track moving face-like patterns with their eyes than other patterns of comparable complexity or symmetry. This finding implies that, prior to virtually any opportunity to learn, the human brain is equipped with a general representation of the appearance of the human face. Another demonstration of innateness in person processing comes from the study of a boy who sustained visual cortical damage, including damage to the fusiform face area, in his first day of postnatal life (Farah et al. 2000). Despite his relatively preserved ability to recognize non-face objects, he never acquired the ability to recognize faces. In other words, a certain region of cortex is destined for face recognition as early as age 1 day, and other regions, which are capable of recognizing inanimate objects, cannot take over this function. This striking absence of plasticity implies that the category of human face, as well as its representation by specific brain tissue, is determined essentially at birth.

Studies with older infants confirm that we distinguish between persons and non-persons, or more accurately between entities that do and do not possess the trigger features for the person network, as early as age 3 months. Moving shapes for which the motions are mutually contingent attract the attention of 3 month-old infants more effectively than shapes moving in non-contingent ways (Rochat et al. 1997). A study of 12 month-old infants found that they, like adults, tend to follow the "gaze" of football-shaped objects (i.e. look where the object seems to be looking) if the object has been seen moving in a contingent manner or if it has eye spots (Johnson 2003).

Infants also implicitly attribute intentions to the behavior of persons at an early age. For example, Woodward (1998) used a habituation paradigm to probe 5-month-old infants' representations of two kinds of events, a person reaching around a barrier to retrieve an object and a mechanical "arm" doing the same thing. The barrier was then removed and one of two things happened next: the reacher (person or machine) reached again for the object with either the same roundabout trajectory or reached for it directly. Infants who saw the machine looked longer (evincing surprise) when it changed its trajectory from roundabout to direct, but those who saw the person looked longer when the trajectory through space was the same as before. This implies that the infants' initial representation of the machine's action concerned its physical motion through space, so that a similar motion was less surprising, whereas their initial representation of the person's action concerned his or her intention to pick up the object, so that a direct retrieval was less surprising.

At the same age, infants grasp certain principles of the physical behavior of objects, including the need to traverse a continuous trajectory through space in going from one point to another, but seem to think of people as exempt from at least certain constraints of physical objects (Kulhmeier et al. 2003). The authors interpret this as evidence for a distinction in the infant's mind between persons and things. Furthermore, because humans do traverse continuous paths through space, this finding was clearly not a learned feature of persons. The authors suggest that it reflects the child's assumption that the important part of a person is the nonmaterial part and the resultant difficulty of thinking of people as physical objects (Bloom 2004).

Another source of evidence for the innate nature of person representation in the brain comes from the study of individuals with autism and autistic spectrum disorders. Autism is a complex condition with a number of cognitive and affective components, but the core feature that distinguishes it from other developmental disorders is abnormal interpersonal behavior. There is a substantial genetic component to autism (Piven 1997) and the social behaviors typical of autism (Ronald et al. 2005). From infancy on, autistic individuals show an unusually low level of interest in other people, generally preferring to interact with inanimate objects. A retrospective study of home movies of first-year birthday parties showed that this tendency was apparent well before the child was diagnosed (Osterling et al. 2002). Autistic children are sometimes described as "treating people like objects," for example attempting to climb a conveniently located adult to get to a toy on a shelf. As adults, autistic individuals have difficulty anticipating the reactions of other people and understanding why others behave as they do. Autistic persons have difficulty with tasks that require representing the mental states of others, for example, understanding the first type of story quoted earlier (Happé et al. 1994).

Functional neuroimaging studies show that the brain regions normally activated for person representation are not activated in autistic participants (see Pelphrey et al). For example, autistic participants do not show increased activation in the person network when viewing the kinds of animated shapes that evoke person-related cognition and neural activity in normal participants (Castelli et al. 2002; Schultz et al. 2003). These participants also tend not to show activation in the fusiform face area when viewing human faces (Critchley et al. 2000; Schultz et al. 2000) and do not activate medial prefrontal cortex when reading stories involving other people's mental states (Happe et al. 1996). These studies suggest that the development of the person network is partly genetically determined, and that autism represents an abnormality in this process.

In sum, we come into the world with a brain system genetically preprogrammed to represent persons as distinct from other kinds of objects in the world. This system is surprisingly autonomous, in the sense that it is triggered by certain stimuli and can be difficult to suppress. It becomes active even when we know that the triggering stimulus is not a person, that is, when other parts of our brain represent the information that the stimulus is not a person, but an unrealistic drawing of a person or even a geometric shape. Indeed, it becomes active in the presence of triggering stimulus features even when irrelevant or downright counterproductive.

CONCLUSIONS

Despite our intuitions that both plants and persons are "out there," in some similar sense of being natural kinds in the world, there are important differences between the two types of category. Science has found an objective basis for the distinction we make intuitively between plants and other multi-cellular organisms, but it has yet to identify useful criteria for personhood. We suggest that this is because the category "plant" has a kind of objective reality that the category "person" does not. In the previous section we summarized evidence that the human brain is born equipped to treat certain types of stimuli—those with such trigger features as a human-like face or body or patterns of movement—in a special way. We perceive them and reason about them using a separate brain system, and do so innately, automatically, and irrepressibly. Our sense that the world contains two fundamentally different categories of things, persons and non-persons, may be a result of the periodic activation of this person network by certain stimuli rather than any fundamental distinction between the stimuli that do and do not tend to trigger it.

Of course, there must be some set of attributes in the world that determine whether or not the person network is triggered. Does that not imply that persons are "in the world" after all? To answer this question, let us consider the relations between mental representation and reality for three categories: persons, plants, and phlogiston. *Phlogiston* is the name of a fluid that 17th and 18th century scientists believed was contained in combustible substances, and that 20th and 21st century philosophers have used to illustrate a point about theory change and word meaning. Combustion was thought to be a process by which phlogiston left the burning substance and was absorbed by the air. In conjunction with some other reasonable assumptions, the phlogiston theory was able to explain a number of different aspects of combustion. For example, the extinction of fires by limiting the air supply could be explained in terms of the air becoming saturated with phlogiston. We now understand that there is no such thing as phlogiston, and that combustion is part of a larger category of phenomena consisting of oxidation. However, when phlogiston theorists perceived burning, their representations of phlogiston became active. These early scientists did not randomly or arbitrarily project a concept of phlogiston onto the world; there was of course some category of events in the world that corresponded in a systematic way to their representation of phlogiston. However, this category was based on relatively superficial perceptual features of the world (e.g., flames) combined in certain ways dictated by their theory, and did not capture any of the deeper or more explanatory structure of nature. The point of this example is that mental representations can exist and be activated by stimuli in systematic ways without picking out fundamental categories of the natural world.

We do not believe that personhood is like phlogiston. Our evolved person representations are probably not as thoroughly wrong as the phlogiston theorists' representations of oxidation. Clearly some things in the world have minds much like our own, and other things do not have minds. There are also different degrees of mindedness, however, and perhaps even different kinds of minds (e.g., Brooks 2002; Edelman et al. 2005). Furthermore, our intuitions about who or what has a mind are partly under the control of superficial and potentially misleading trigger features such as eyes and faces. In this sense, our person representations do

not reflect reality as accurately as our plant representations. We suggest that two features of person representation in the brain underlie this discrepancy.

The first relevant feature of the person network in the brain is its separateness from the systems representing other things. We suggest that this feature is responsible for the illusion that persons and non-persons are fundamentally different kinds of things in the world, despite our inability to draw a principled line between them. This illusion may come from the operation of two separate and incommensurate systems of representation in the brain for persons and for things in general, in contrast to a common distributed representation. Within a unitary distributed representation of color, for example, one could represent a shade that is red or orange, and if both are active, one would automatically be representing reddish orange (see O'Reilly and Munakata 2000). Within a unitary distributed representation for shape with representations of bowl-like and cuplike forms, the simultaneous activation of both would represent an object with an in-between shape, a kind of large, wide cup. But what if red or "bowlness" are also represented by a separate system from other colors and other shapes? Then regardless of how much orange or "cupness" is being registered elsewhere in the brain, the sight of reddish orange or a large, wide cup will result in a representation of red or a bowl. These representations of red or bowl shape may be weaker than those engendered by a true red or a prototypical bowl shape, but they will nevertheless be weakly red as opposed to reddish orange, or weakly bowl-like as opposed to bowlish-cuplike.

Someone perceiving the world with such a system of representation would perceive both the continuities among colors and shapes, but also the existence of a divide between red things and non-red things, bowls and non-bowls. Such a person might say "I can't find a sensible place to draw a line between red and reddish-orange things, but it seems clear to me that some things have redness and some do not. Things may vary in how much redness they have, but by having redness they are fundamentally different from other things." Substitute the person system for the red system, and one gets the very intuition that has posed such a problem in philosophy and bioethics. This intuition could be expressed thus: "People, animals, and even computers may have varying amounts of intelligence, communication ability, and self-awareness. I can't find a sensible place to draw a line across the potential continuum of states linking, say, a healthy human and one in a vegetative state or linking a current-day computer and one endowed with humanlike intelligence. Nevertheless, I have the sense that some beings have personhood and others do not."

The second relevant feature of the person network is its autonomy, its tendency to become activated by certain triggering stimuli (e.g., faces and contingent behavior) whether or not we believe there is actually a person there. Even if persons were like plants, and there were a clear objective basis for separating persons and non-persons, the relentless projection of personhood on the basis of fragmentary cues would lead to error and confusion on its own. For example, the human face is a powerful trigger cue that activates the whole person network, and this may be what makes it hard for many of us to dismiss the personhood of a vegetative patient or a fetus. If we had a plant network and it functioned similarly, we might feel the urge to sniff the flowers on a friend's Hawaiian print shirt or water carpets that are green.

Why would such a misleading system for person representation have evolved? The answer most likely concerns the intensely social nature of our species and also perhaps the rarity of ambiguous cases of personhood in our evolutionary history. Like other social species, our individual survival depends on relating successfully to our conspecifics. More for us than for other species, this requires understanding the immensely complex behaviors that result from their beliefs, motivations, and personalities. As the anthropologist Guthrie (1995) has observed, in discussing religious belief systems, the cost of attributing intentionality to some non-intentional systems may be less than the cost of failing to adopt the intentional stance toward some systems that are intentional. In other words, it may have been adaptive to err on the side of activating the personhood network too often.

Furthermore, the personhood network is an adaptation to an earlier world, which contained fewer ambiguous cases of personhood. Sonograms did not show us our fetuses; people did not live long enough to develop Alzheimer's disease, and vegetative states were fatal. It is interesting that infants and young children may be the one class of ambiguous cases that our ancestors did encounter on a regular basis, and for these cases it would be adaptive to attribute personhood even in the absence of intelligence and self-awareness. Protohumans who accurately judged their offspring to be lacking in the various traits associated with personhood and accordingly treated them as non-persons would not have many surviving descendents!

If our analysis is correct, it suggests that personhood is a kind of illusion. Like visual illusions, it is the result of brain mechanisms that represent the world nonveridically under certain circumstances. Also like visual illusions, it is stubborn. Take the Hermann grid illusion, for example, in which a grid of white lines on a black background seems to have ghostly grey spots at the lines' intersections (see http://www.yorku.ca/eye/hermann.htm). We know that these spots are illusory, and that they result from interactions between the antagonistic center and surround compartments of the receptive fields of visual neurons; however, this knowledge does not make the spots go away! Similarly, knowing about the person network does not eliminate the sense that moving Heider and Simmel shapes have intentions.

The result of this analysis could be considered nihilistic. It does undercut ethical systems based on personhood, and in particular suggests that difficult ethical issues should not be approached with the strategy of determining whether or not the parties involved are persons. If personhood is not really in the world, then there is no fact of the matter

concerning the status of a given being as a person or not, and there is no point to the philosophical or bioethical program of seeking objective criteria for personhood more generally because there are none.

Where does this leave us? The answer is different for ethics, as a discipline, and for the everyday moral behavior of individuals. For ethics, the only alternative we can see is a shift to a more utilitarian approach. Rather than ask whether someone or something is a person, we should ask how much capacity exists for enjoying the kinds of psychological traits previously discussed (e.g., intelligence, self-awareness) and what are the consequent interests of that being. Of course, this view requires deciding how these traits should be defined and ranked in importance and whether to consider a being's potential, or only actual, status. In other words, many similar problems arise as in discussions of criteria for personhood. However, having understood the need to set aside intuitions about personhood and having avoided the distraction of seeking criteria for personhood, we can work more productively on assessing and protecting the interests of all.

In contrast, as individuals whose behavior includes countless implicit moral decisions each day, it matters little whether personhood is illusion or reality. We cannot reprogram ourselves to stop thinking in terms of persons, nor would we want to. It is thanks to this stubborn illusion that we persist in talking to our babies, who cannot understand what we are saying, but who clearly benefit from the social and linguistic stimulation. It is thanks to the personhood network's hair trigger that we slam on the brakes at the first glimpse of a human form in the road, rather than wait until our conscious mind has arrived at the belief that there is someone there. Although the concept of personhood may be bad metaphysics and better suited to an earlier world, even today it serves us well. In this respect, we are like the guy in the joke with the brother who thinks he's a chicken. When asked why he does not take his brother to a psychiatrist to be cured, he answers: "Because I need the eggs."

REFERENCES

Adolphs, R. 2003. Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience* 4: 165–78.

Adolphs, R., F. Gosselin, T. W. Buchanan, D. Tranel, P. Schyns, and A. R. Damasio. 2005. A mechanism for impaired fear recognition after amygdala damage. *Nature* 433: 68–72.

Beauchamp, T. L., and J. F. Childress. 2001. *Principles of biomedical ethics*. 5th ed. New York, NY: Oxford University Press.

Bloom, P. 2004. *Descartes' baby: How the science of child development explains what makes us human*. New York, NY: Basic Books.

Brooks, R. A. 2002. *Flesh and machines*. New York, NY: Pantheon Books.

Brothers, L. 1990. The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience* 1: 27–51.

Castelli, F., C. D. Frith, F. Happé, and U. Frith. 2002. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* 125: 1839–1849.

Castelli, F., F. Happé, U. Frith, and C. Frith. 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* 12: 314–325.

Critchley, H. D., E. M. Daly, E. T. Bullmore, et al. 2000. The functional neuroanatomy of social behaviour: changes in cerebral blood flow when people with autistic disorder process facial expressions. *Brain* 123: 2203–2212.

Dennett, D. 1978. *Brainstorms: Philosophical essays on mind and psychology*. Cambridge, MA: The MIT Press.

Downing, P. E., D. Bray, J. Rogers, and C. Childs. 2004. Bodies capture attention when nothing is expected. *Cognition* 93: B27–B38.

Downing, P. E., Y. Jiang, M. Shuman, and N. Kanwisher. 2001. A cortical area selective for visual processing of the human body. *Science* 293:2470–2473.

Edelman, D. B., B. J. Baars, and A. K. Seth. 2005. Identifying hallmarks of consciousness in non-mammalian species. *Consciousness and Cognition* 14: 169–187.

Engelhardt, H. T. J. 1986. *The foundations of bioethics*. New York, NY: Oxford University Press.

Farah, M. J. 2000. *The cognitive neuroscience of vision*. Oxford, UK: Blackwell Publishers.

Farah, M. J. 2004. *Visual Agnosia. 2nd ed.* Cambridge, MA: The MIT Press.

Farah, M. J., K. L. Levinson, and K. L. Klein. 1995. Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia* 33: 661–74.

Farah, M. J., C. Rabinowitz, G. Quinn, and G. Liu. 2000. Early commitment of the neural substrates of face recognition. *Cognitive Neuropsychology* 17: 117–123.

Farah, M.J., K.D. Wilson, H.M. Drain, and J.R. Tanaka. 1995. The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research* 35: 2089–2093.

Feinberg, J. 1980. Abortion. In *Matters of life and death*, ed. T. Regan, 188–189. Philadelphia, PA: Temple University Press.

Feinberg, T. E., R. J. Schindler, E. Ochoa, P. C. Kwan, and M. J. Farah. 1994. Associative visual agnosia and alexia without prosopagnosia. *Cortex* 30: 395–411.

Fellows, L. K., and M. J. Farah. 2005. Dissociable elements of human foresight: A role for the ventromedial frontal lobes in framing the future, but not thinking about future rewards. *Neuropsychologia* 43: 1214–1221.

Fletcher, J. 1979. *Humanhood: Essays in biomedical ethics*. Buffalo, NY: Prometheus Books.

Fletcher, P. C., F. Happe, U. Frith, S. C. Baker, R. J. Dolan, R. S. Frackowiak, and C. D. Frith. 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition* 57: 109–128.

Gallagher, H., F. Happe, N. Brunswick, P. Fletcher, U. Frith, and C. Frith. 2000. Reading the mind in cartoons and stories: An fMRI study of 'theory of mind'. *Neuropsychologia* 38: 11–21.

Gallagher, H. L., A. I. Jack, A. Roepstorff, and C. D. Frith. 2002. Imaging the attentional stance in a competitive game. *NeuroImage* 16: 814–821.

Gertler, G. B. 1986. Brain birth: A proposal for defining when a fetus is entitled to human life status. *Southern California Law Review* 59: 1061–1078.

Goel, V., J. Grafman, N. Sadato, and M. Hallett. 1995. Modeling other minds. *Neuroreport* 6: 1741–1746.

Green, R. M. 2002. Part III: Determining moral status. *American Journal of Bioethics* 2: 20–30.

Grossman, E., M. Donnelly, R. Price, et al. 2000. Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience* 12: 711–720.

Guthrie, S. E. 1995. *Faces in the clouds: A new theory of religion.* Oxford, UK: Oxford University Press.

Happé, F., S. Ehlers, P. Fletcher, U. Frith, M. Johansson, C. Gillberg, R. Dolan, R. Frackowiak, and C. Frith. 1996. 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. *Neuroreport* 8: 197–201.

Happé, F. G. 1994. An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders* 24: 129–154.

Haring, B. 1972. *Medical ethics.* Slough, MN: St. Paul Publications.

Heberlein, A. S., and R. Adolphs. 2004. Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proceedings of the National Academy of Science USA* 101: 7487–7491.

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

Johnson, M. H., S. Dziurawiec, H. Ellis, and J. Morton. 1991. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40: 1–19.

Johnson, S. C. 2003. Detecting agents. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences* 358: 549–559.

Jones, D. G. 2004. The emergence of persons. In *From cells to souls — and beyond*, ed. M. Jeeves, 11–33. Grand Rapids, SD: William B. Eerdmans.

Jones, G. 1989. Brain birth and personal identity. *Journal of Medical Ethics* 15: 173–178.

Kant, I. 1948. Groundwork of the metaphysics of morals. In *The moral law: Kant's groundwork of the metaphysics of morals*, ed. H. J. Paton, X–XX. London, UK: Hutchinson.

Kanwisher, N., J. McDermott, and M. M. Chun. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* 17: 4302–4311.

Locke, J. 1997. *An essay concerning human understanding.* Harmondsworth, UK: Penguin Books.

Martin, A., and J. Weisberg. 2003. Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology* 20: 575–587.

Martin, R. C. 2003. Language processing: Functional organization and neuroanatomical basis. *Annual Review of Psychology* 54: 55–89.

McMahan, J. 1995. The metaphysics of brain death. *Bioethics* 9: 91–126.

McNeil, J. E., and E. K. Warrington. 1993. Prosopagnosia: A face-specific disorder. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 46A: 1–10.

Miller, E. K., and J. D. Cohen. 2001. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 24: 167–202.

Mitchell, J. P., T. F. Heatherton, and C. N. Macrae. 2002. Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences U S A* 99: 15238–15243.

Mitchell, J. P., C. Neil Macrae, and M. R. Banaji. 2005. Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *NeuroImage* 26: 251–257.

Moussa, M., and T. A. Shannon. 1992. The search for the new pineal gland. Brain life and personhood. *Hastings Center Report* 22: 30–37.

O'Reilly, R. C., and Y. Munakata. 2000. *Computational explorations in cognitive neuroscience.* Cambridge, MA: The MIT Press.

Osterling, J. A., G. Dawson, and J. A. Munson. 2002. Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology* 14: 239–251.

Paulesu, E., J. Harrison, S. Baron-Cohen, J. D. Watson, L. Goldstein, J. Heather, R. S. Frackowiak, and C. D. Firth. 1995. The physiology of coloured hearing. A PET activation study of colour-word synaesthesia. *Brain* 118 (Pt 3): 661–676.

Peelen, M. V., and P. E. Downing. 2005. Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology* 93: 603–608.

Pelphrey, K., R. Adolphs, and J. P. Morris. 2004. Neuroanatomical substrates of social cognition dysfunction in autism. *Mental Retardation and Developmental Disabilities Research Review* 10: 259–271.

Phillips, M. L., W. C. Drevets, S. L. Rauch, and R. Lane. 2003. Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological Psychiatry* 54: 504–514.

Piven, J. 1997. The biological basis of autism. *Current Opinion in Neurobiology* 7: 708–712.

Rochat, P., R. Morgan, and M. Carpenter. 1997. Young infants' sensitivity to movement information specifying social causality. *Cognitive Development* 12:537–561.

Ronald, A., F. Happe, and R. Plomin. 2005. The genetic relationship between individual differences in social and nonsocial behaviours characteristic of autism. *Developmental Science* 8: 444–458.

Rorty, A. O. 1988. *Mind in action: Essays in the philosophy of mind.* Boston, MA: Beacon Press.

Saxe, R., and A. Wexler. 2005. Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia* 43: 1391–1399.

Saxe, R., D. K. Xiao, G. Kovacs, D. I. Perrett, and N. Kanwisher. 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42: 1435–1446.

Scholl, B. J., and P. D. Tremoulet. 2000. Perceptual causality and animacy. *Trends in Cognitive Sciences* 4: 299–309.

Schultz, R. T., I. Gauthier, A. Klin, R. K. Fulbright, A. W. Anderson, F. Volkmar, P. Skudlarski, C. Lacadie, D. J. Cohen, and J. C. Gore. 2000. Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger syndrome. *Archives of General Psychiatry* 57: 331–340.

Schultz, R. T., D. J. Grelotti, A. Klin, J. Kleinman, C. Van der Gaag, R. Mavois, and P. Skudlarski. 2003. The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society of London B. Biological Sciences* 358: 415–427.

Singer, P. 1994. *Rethinking life and death: The collapse of our traditional ethics.* Oxford, UK: Oxford University Press.

Skuse, D., J. Morris, and K. Lawrence. 2003. The amygdala and development of the social brain. *Annals of the New York Academy of Science* 1008: 91–101.

Squire, L. 1992. Memory and the hippocampus: A synthesis of findings from rats, monkeys and humans. *Psychological Review* 99: 195–231.

Tarr, M. J., and I. Gauthier. 2000. FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience* 3: 764–769.

Tooley, M. 1972. Abortion and infanticide. *Philosophy and Public Affairs* 2: 37–65.

Vuilleumier, P. 2000. Faces call for attention: Evidence from patients with visual extinction. *Neuropsychologia* 38: 693–700.

Wise, S. 2002. 'Practical autonomy' entitles some animals to rights. *Nature* 416: 785.

Woodward, A. L. 1998. Infants selectively encode the goal object of an actor's reach. *Cognition* 69: 1–34.

Wright, C. I., B. Martis, L. M. Shin, H. Fischer, and S. L. Rauch. 2002. Enhanced amygdala responses to emotional versus neutral schematic facial expressions. *Neuroreport* 13: 785–790.