mistaken identifications of items, or individuals, as persons. It appears that we are inclined to overinterpret the personhood of inanimate systems and simple organisms through a process of inappropriate personification. This neurological evidence suggests that much animistic superstition can be understood as inappropriate prefrontal cortical activation. Whenever we curse our computers for crashing or our cars for not starting, we are very likely in the grip of our easily and inappropriately triggered "social brain." At a rational level we are very likely to be perfectly aware of the inappropriate projective personification that is taking place, but the impulse is irresistible—perhaps because we are, as it were, hard-wired for superstition.

The fact that our person-recognizing system may be inappropriately triggered in this way in no way vitiates the concept of a person. It just shows that we are particularly prone to what statisticians call *type 1 errors*, that is, to accepting false-positives. Evolution has wisely calibrated the person-recognizing neurological networks of social animals that lived in small groups to err in this direction. Almost certainly it was better for us, as social beings, to mistakenly treat non-persons as persons than to commit the opposite error. Autistic individuals, in contrast, are socially disadvantaged by being inclined to make the opposite (false-negative) error, that is, they fail to identify persons as persons.

The mistaken identifications that result from the inappropriate activation of our hair-trigger person-recognition systems nevertheless generate a lot of appropriate and indeed vitally important adaptive behavior. As Farah and Heberlein (2007) point out, treating babies as persons may be an epistemological (as well as a metaphysical) error, but it is certainly a socially productive one from the point of view of human societies.

None of these interesting observations establish that the concept of a person is useless for ethics. Nor do they compel us to adopt a utilitarian approach to ethics. There may be good arguments in favour of utilitarianism as an ethical theory, but this is not one of them. In any case, utilitarianism can only be developed on the basis of an understanding of pleasures or preferences being satisfied (or thwarted), and these basic psychological notions face exactly the same problems of identifications as the notion of a person.

Where does this leave us? It does not compel us to reject ethical systems that are based on personhood. But we probably do need to acknowledge that ethical systems that presuppose a concept of personhood are likely to be too generously applied because human beings are unreliable person-detectors. This does not mean that there are no persons of whom we should be morally considerate. Rather it suggests that we are liable to be morally considerate to a fault, precisely because we will inevitably draw the boundaries of personhood too generously. There is no stubborn illusion that there are persons—only a persistent (but potentially correctable) bias to be over-generous in collecting things or individual beings under the protective moral umbrella of a "person."

What role does neuroscience have in helping to resolve ethical dilemmas? Certainly neuroscience has helped us to better understand the way in which we make systematic overestimations of personhood and has thereby helped us to understand how our moral intuitions about personhood can be seriously mistaken. Neuroscience has also helped us to develop the notion of "brain death" (the death of a person) and this conception has been instrumental in providing a better understanding of end-of-life ethical issues and in helping us to resolve these dilemmas. The potential to become persons (as in babies), the ability to recover personhood (e.g., from neurological injury) and the loss of personhood (e.g., neurological degeneration) are vital points on our moral compass, and they are all issues about which neuroscience has profoundly deepened our understanding.

REFERENCES

Churchland, P. S. 2007. The necessary-and-sufficient boondoggle. *American Journal of Bioethics (AJOB–Neuroscience)* 7(1): 54–55.

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

Neuroscience and Metaphysics (Redux)

Chris Buford, University of Redlands
Fritz Allhoff, Western Michigan University

In earlier work, we challenged a contention made by Illes and Racine (2005) that developments in neuroscience have significant metaphysical implications (Buford and Allhoff 2005). Farah and Heberlein (2007) have advanced a similar

Address correspondence to Fritz Allhoff, Department of Philosophy, Western Michigan University, 1903 W. Michigan Ave., Kalamazoo, MI 49008. E-mail: fritz.allhoff@wmich.edu or Chris Buford, Department of Philosophy, University of Redlands, Redlands, CA 92373. E-mail: chris_buford@redlands.edu

contention, although pertaining to different metaphysical targets. Before reviewing their scientifically informed claims, we want to raise some challenges to their philosophical ones.

In their article, Farah and Heberlein (2007) argue that the failure to discover "non-specific and non-arbitrary" necessary and sufficient conditions for personhood is best explained by the fact that persons do not exist.[1] Our first concern is their apparent contention that the absence of necessary and sufficient conditions for some concept implies that the concept is *illusory* (i.e., that the world does not contain any entities to which the concept applies). Owing largely to the work of Ludwig Wittgenstein (1999, 65), there is close to a near-philosophical consensus that necessary and sufficient conditions cannot be provided for a wide range of concepts, yet we certainly do not want to jettison entities or relations putatively referred to by these concepts—such as *minds*, *tables* or *knowledge*—from our ontologies. If the absence of these joint conditions is merely necessary, rather than sufficient, for a concept to be illusory, then Farah and Heberlein still owe us an account of what other features are needed for such a rejection.

Farah and Heberlein (2007) do seem to provide another feature that might be doing that work: the etiology of the belief in question. This line would hold that if we possess a concept for some particular reason (which has yet to be elucidated), then we might be justified in being more skeptical about the ontological status of kinds or objects corresponding to the concept in question. Farah and Heberlein offer the following:

> Perhaps this intuition [that persons exist] 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 . . . neuroscience may show us that personhood is illusory (2007, 37).

The authors do present a wealth of scientific data supporting the claim that various neurological structures correlated with social interaction are activated at an extremely early age and that those structures activate independently of the individual's beliefs about whether or not a person is genuinely present. *Innateness* of some variety would surely be an explanation of such findings. However, there seems to be a large gap between such results and the existence of a complete *explanation* of our intuition that persons do exist. Imagine living in a world without any other persons. Instead, your world was populated with myriad inanimate objects that moved around in patterns stimulating the neural structures that would be stimulated if you were to interact with real persons. Would you have the intuition that persons (other than yourself) exist? Surely not. Thus, even if the innateness hypothesis is true, it alone cannot explain the intuition that persons exist. The additional explanatory work is supplied by our interaction with individuals having many of the traits thought to be possessed by persons.

Furthermore, even if the intuition stems from the fact that the concept of personhood "structures our experience of the world from the outset," why ought we to think that neuroscience has given us good reason to think that persons do not exist? Farah and Heberlein explain:

> 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 (2007, 37).

Thus, the best explanation of such a vocabulary, even if it were innate in a certain sense, would be the survival value of recognizing persons.

Farah and Heberlein (2007) also attempt to explain the intuition by concentrating on the isolation and autonomy of the brain's "person network." They argue that the person network's isolation gives rise to the illusory belief that persons are fundamentally different from non-persons despite the absence of an agreed line of separation. A color-perception analogy is pursued to show how isolation of neural networks can lead to error:

> Someone perceiving the world with such a system of representation would perceive both the continuities. . . between red things and nonred things . . . 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 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 (Farah and Heberlein 2007, 37).

However, is something wrong with this line of reasoning? Things that are red are fundamentally different than non-red things, they are red, and there is no sharp line between red things and non-red things. Yet, red things do exist. If Farah and Heberlein are to show that isolation can explain our intuition that persons do not exist, then we need evidence that such isolation is likely to lead to large-scale misrepresentation. The example cited falls short of accomplishing this goal. We should also note that in the case of vagueness with respect to when an object of one sort becomes an object of another sort, it does not follow that objects of neither sort exist; to think otherwise is to reject the intuitively plausible notion of metaphysical vagueness.

It might be that we are supposed to read "fundamental" as requiring that a fundamental difference can only be a difference in essential properties. In this case, holding red things to be fundamentally different from non-red things is probably an error. Even if this interpretation is correct, it is not clear that this is sufficient to show that persons do not exist. It may be true that many of us are mistaken,

[1] We see the authors as concerned with the necessary and sufficient conditions of being a person at a specific time. This question is prior to the traditional question of personal identity that asks for the necessary and sufficient of identity across time between persons. If there are no persons, then there are no necessary and sufficient conditions for personal identity across time.

that persons are not a fundamental kind of thing (i.e., that something can be a person at one time, yet fail to be a person and still exist at another time). However, this certainly does not prove that persons do not exist, nor does it show that the concept of personhood is irrelevant to morality. You and I might be identical to human animals, animals that at some points in our lives possess the requisite cognitive abilities to be classified as *persons*. This metaphysical picture still seems compatible with the importance of personhood to moral matters.

The fact that our person network evidences a "tendency to become activated by certain stimuli . . . whether or not we believe there is actually a person there" (Farah and Heberlein 2007, 37) also fails to adequately explain the non-existence of persons. Farah and Heberlein are correct to point out that "the cost of attributing intentionality to some non-intentional systems may be less than the cost" of non-attribution and that the "personhood network is an adaptation to an earlier world, which contained fewer ambiguous cases of personhood" (2007, 37). Both considerations hardly support the massive error theory posited by Farah and Heberlein for the same sort of reasoning would seem to support the non-existence of thinking intelligent beings.

More generally, though, there seems to be some basic flaw in their reasoning: just because some belief has some sort of evolutionary explanation, it hardly follows that the object of that belief does not exist. For example, we have evolved to believe in the presence of tigers and cliffs, and it is obviously implausible to say that tigers and cliffs do not exist. As we noted previously, their argument seems to be that philosophical trouble (regarding conceptual analysis) *plus* a certain etiology for the belief is grounds for divesting ourselves of some relevant concept (in their case, personhood), but this just cannot be right. The concept of a tiger and a cliff will also resist submission to any sort of naïve conceptual analysis, and we can attribute our beliefs thereof to some sort of evolutionarily enabled perceptual faculties, yet it does *not* follow that tigers and cliffs do not exist. Granting their contention, then, that 'personhood' (or 'persons') has the same features, it would not follow that we should reject it either.

It might be the case that their two conditions for rejection of a concept are necessary, rather than sufficient. In other words, it *could* be the case that we want to retain tigers and cliffs in our ontologies while rejecting persons, despite the fact that all have the previously mentioned features. But then we would need some further differentia between those classes, and Farah and Heberlein (2007) do not provide any. Maybe there is some difference between perceptual beliefs (e.g., tigers and cliffs) as against more "abstract" beliefs (e.g., personhood), although some principled account would have to be offered therein. At any rate, denying the sufficiency of their proposed conditions (through the counterexamples suggested previously) is enough to show that their conclusion (*viz*., that we should reject 'personhood') does not follow. In general, and as we have previously argued (Buford and Allhoff 2005), we simply do not think that neuroscience is going to have many philosophical implications, particularly regarding metaphysics.

REFERENCES

Buford, C., and F. Allhoff. 2005. Neuroscience and metaphysics. *American Journal of Bioethics* 5(2): 34–36.

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

Illes, J., and E. Racine. 2005. Imaging or imagining?: A neuroethics challenge informed by genetics. *American Journal of Bioethics* 5(2): 5–18.

Wittgenstein, L. 1999. *Philosophical investigations*, 3rd edition. New York, NY: Prentice Hall.

Personhood: Elusive But Not Illusory

John Banja, Emory University

Richard Rorty (1979, 1991) has taunted philosophers for decades about their Platonic neurosis, that is, their penchant for conjuring up a hyper-real world of essences that exists independently of human cultures and that provides a transcendental grounding for the words we use and the objects they represent (Hance 1995). Socrates is the disease's most famous representative, afflicting philosophers for millennia by asking Euthyphro in the dialogue of that name:

> Is not the holy always one and the same thing in every action, and, again, is not the unholy always opposite to the holy? . . . And as unholiness, does it not always have its one essential

Address correspondence to John Banja, Ethics Center, Emory University, 1462 Clifton Road, Suite 302, Atlanta, GA 30322. E-mail: jbanja@emory.edu