



University of Pennsylvania  
**ScholarlyCommons**

---

Neuroethics Publications

Center for Neuroscience & Society

---

June 2007

## Response to Open Peer Commentaries on "Personhood and Neuroscience: Naturalizing or Nihilating?": Getting Personal

Martha J. Farah  
*University of Pennsylvania*, [mfarah@psych.upenn.edu](mailto:mfarah@psych.upenn.edu)

Andrea S. Heberlein  
*Harvard University*

Follow this and additional works at: [https://repository.upenn.edu/neuroethics\\_pubs](https://repository.upenn.edu/neuroethics_pubs)

---

### Recommended Citation

Farah, M. J., & Heberlein, A. S. (2007). Response to Open Peer Commentaries on "Personhood and Neuroscience: Naturalizing or Nihilating?": Getting Personal. Retrieved from [https://repository.upenn.edu/neuroethics\\_pubs/36](https://repository.upenn.edu/neuroethics_pubs/36)

Postprint version. Published in *The American Journal of Bioethics*, Volume 7, Issue 1, January 2007, pages 37-48.  
Publisher URL: <http://dx.doi.org/10.1080/15265160601150352>

This paper is posted at ScholarlyCommons. [https://repository.upenn.edu/neuroethics\\_pubs/36](https://repository.upenn.edu/neuroethics_pubs/36)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

## Response to Open Peer Commentaries on "Personhood and Neuroscience: Naturalizing or Nihilating?": Getting Personal

### 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 *The American Journal of Bioethics*, Volume 7, Issue 1, January 2007, pages 37-48.

Publisher URL: <http://dx.doi.org/10.1080/15265160601150352>

This article was downloaded by:[University of Pennsylvania]  
[University of Pennsylvania]

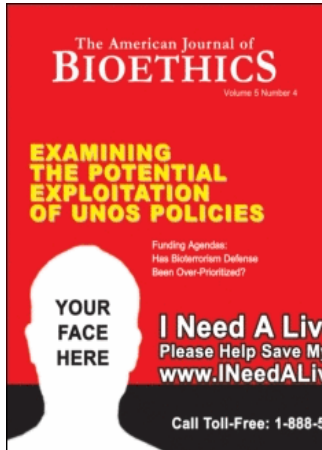
On: 6 June 2007

Access Details: [subscription number 769795022]

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954

Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## The American Journal of Bioethics

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713606739>

### Response to Open Peer Commentaries on "Personhood and Neuroscience: Naturalizing or Nihilating?": Getting Personal

To cite this Article: Farah, Martha J. and Heberlein, Andrea S. , 'Response to Open Peer Commentaries on "Personhood and Neuroscience: Naturalizing or Nihilating?": Getting Personal', The American Journal of Bioethics, 7:1, W1 - W4

To link to this article: DOI: 10.1080/15265160601150352

URL: <http://dx.doi.org/10.1080/15265160601150352>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

© Taylor and Francis 2007

Correspondence

# Response to Open Peer Commentaries on “Personhood and Neuroscience: Naturalizing or Nihilating?”: Getting Personal

Martha J. Farah and Andrea S. Heberlein, Center for Cognitive Neuroscience, University of Pennsylvania

Why do we have such strong intuitions that persons—as opposed to merely humans, thinking things, or sentient beings—exist? Why can’t we identify the traits that determine personhood, or agree on which entities are persons? These are the main questions that our Target Article addresses. We suggest that the answer involves innate mechanisms that served us (or our genes) well enough in ancient times, but which create confusion today. Indeed, we suggest that the very intuition that there is a category of things in the world corresponding to persons is confused. Furthermore, our attempts to identify traits that confer moral standing, by virtue of personhood, suffer from the influence of our evolved intuitions on our reasoning about morally relevant psychological and neurological characteristics.

The first part of our argument is that personhood has not yet been satisfactorily defined, raising the possibility that it does not correspond to a real category of things in the world. Many of the commentators (Banja 2007; Buford and Allhoff 2007; Churchland 2007; Grey, Hall and Carter 2007; Nelson 2007; Perring 2007; Sagoff 2007) address this part of the argument. To summarize their critique (with blithe disregard of the nuances that distinguish their positions), they point out that this is nothing new, that indeed most concepts are fuzzy and/or impossible to define. To this we have two responses. First, the personhood concept suffers from more than just fuzziness; its definitional problems run deeper than the lack of a sharp boundary. Second, the difficulty of defining personhood is only the first half of our argument; to reconstruct the argument as “personhood resists definition therefore there are no persons” is to get it wrong.

The problem of defining personhood is more fundamental than a mere lack of necessary and sufficient criteria. We know that fuzzy boundaries separate mountains and molehills, day and night, etc. The assumption that natural categories have sharp boundaries is a straw man, and it is not part of our argument—not even the first part of our argument. *Simply recognizing that personhood could be a matter of de-*

*gree does not solve the definitional problems we reviewed.* Rosch’s subjects might disagree on whether a lamp is “furniture,” but they agree that it belongs between a sofa and a door-knob in its degree of “furniture-ness”—a pattern of data that is a hallmark of fuzzy categories. In contrast, there is no consensus on the rankings of the “personness” of a patient in a persistent vegetative state (PVS), a fetus and an ape. There are reasonable people who would put PVS patients, late-term fetuses and apes clearly on one side or the other of the fuzzy divide. Corresponding to this lack of consensus on the classification of various entities as persons is the definitional problem on which we focused in the first part of our article, namely the lack of consensus or obvious correct choice concerning which traits (psychological or neurological) determine personhood.

Whereas some of the commentators found the undefinability of personhood so obvious as to be trivial (though they all put it more nicely than that), we note that some of the greatest minds in the history of philosophy and bioethics have sought to define personhood. Furthermore, other commentators seemed to hold out hope for a definition, and indeed some even took a crack at it. For example, Grey, Hall and Carter assert that the psychological characterization of being “rational and self-aware . . . conscious of themselves as agents with a history and the capacity to shape their own future” (2007, 57) is sufficient for personhood. Glannon states that persons are generally taken to be “beings with the capacity for consciousness, the capacity to interact with others, and to conform to social norms when we act” (2007, 57), and adds that neuroscience shows us that “persons are just sets of biological properties generated and sustained by cortical and subcortical structures in the brain” (57). Although these proposals seem like good beginnings, the challenge is to make them specific enough to capture our intuitions about which entities are and are not persons (or for fuzzy boundary lovers, our intuitions about degrees of personhood). For example, without substantial further revision the proposals above would arguably confer personhood on

certain animals and deny it to humans with certain neuropsychiatric conditions. In sum, the claim that personhood has yet to be defined, in either a categorical or graded manner, seems to have struck some as obvious but seems not to have been accepted by others.

In any case, our argument concerning the illusory nature of personhood does not rest on these definitional problems per se. We offered the failures of previous attempts to define or naturalize personhood as a reason to ask ourselves, “should we be assuming that persons really are in the world, and that by studying the world we can determine which entities are persons and which are not?” As just noted, many good minds have assumed this. What we take from the failure of these programs is not that there are no persons—clearly a non sequitur—but that it is *worth considering other reasons for having the intuition of persons other than their actual existence as a natural kind in the world*. A proposal for one such reason is the second part of our argument.

The second part of our argument concerns the source of our intuitions about personhood. Given that defining criteria, even of a graded kind, have yet to be identified for the concept “person,” it is worth inquiring why we have such a strong intuition that the world contains the category persons of course it is possible that our intuitions do reflect empirical generalizations from experience with a world containing persons and non-persons, and that our attempts to characterize personhood explicitly have simply not yet hit on the key traits that correlate with personhood. As stated earlier, our argument is not that the absence of a definition alone implies that personhood is an illusion. However, given the state of play in the attempt to define personhood, it worth asking where else our vivid intuitions—that there are persons might have come from.

As it turns out there is evidence from neuroscience and developmental psychology that these intuitions may come from innate brain mechanisms, separate from those used to learn about other concepts. *Whereas the mental categories day and night, mountain and molehill are learned primarily by induction from experience with the world, we come into the world equipped with the tendency to categorize certain entities as persons and to treat them accordingly*. Things with such trigger features as humanlike faces and/or contingent movement and/or natural language will tend to elicit from us the “intentional stance” (the assumption that the entity is acting according to its own beliefs and desires; Dennett 1978) and a certain moral regard.

Although some commentators questioned our use of the term “illusion” in connection with personhood (Banja 2007; Nelson 2007; Roskies 2007), we stand by it in the sense that the presence of trigger features can cause a host of psychological and moral traits to be projected onto objects in the world, quite independently of what our reason tells us. For example, recall that the presence of eyes on a computer screen increases the generosity with which people play a computer game (Farah and Heberlein 2007, 37). Have a look at videos of the robot Kismet (<http://www.ai.mit.edu/projects/sociable/videos.html>)

and see how hard it is to view the robot as merely a mobile metallic object. This is why we say that personhood is an illusion; just as knowledge of the real nature of a visual stimulus does not vanquish a visual illusion—the moon still looks bigger at the horizon and the intersections of the Herring grid still look dark, even after we have been taught about these illusions—so the knowledge that the eyes on the computer are just paper decals and the knowledge that Kismet is a piece of machinery fail to vanquish our person-related responses to these objects. Kismet still seems sentient and it would feel a bit distressing to smash it with a baseball bat.

Some commentators attempt to justify retaining the concept of a person while acknowledging both the definitional difficulties and the evidence that our thinking about persons is influenced by innate systems rather than induced from experience with the world. The boldest attempts along these lines were by Buford and Allhoff (2007) and by Meghani (2007), who in different words made the point that just because (a) we haven’t been able to identify so far the correlates of a concept in the natural world, and (b) we have innate mechanisms causing us to project that concept onto the world, it does not follow that (c) the concept does not exist in the world. This argument is logically correct but, if we understand it, silly. It seems analogous to the following: If we are watching a movie in Philadelphia, where (a) we have been unable to ascertain that Angelina Jolie and Brad Pitt are in town and (b) we know there is a projector in the theater showing a film starring these actors, nevertheless (c) we cannot rule out that the figures we see at the front of the theater are actually Angelina and Brad acting their movie roles in person in front of the screen.

Roskies’ (2007) attempt to rescue personhood reframes our argument in terms of two systems for thinking about persons, one of which is innate and responsible for the illusory aspects of personhood, and the other of which is more objective (based on psychological traits such as rationality, communication, etc). She suggests we may have gone overboard by calling personhood an illusion; it would be more accurate to say that persons are real but we are prone to misjudgments about them because of that first system. A different point, suggested by Nelson, is that many entities are paradigm instances of persons and that the concept of person works for these, even if it needs clarification for more debatable instances.

The problem with these approaches comes when we try to separate the two versions proposed by each commentator, let us call them “intuitive” and “objective” for Roskies (2007) and “paradigm” and “hard” cases for Nelson (2007). In the first case, the two versions may not be so easily disentangled. The concept “person” comes with heavy baggage from our intuitive ways of responding to humans and other entities, and it is unclear whether we can free ourselves of the grip of, say, faces when we deliberate about the objective status of a candidate person. For this reason, we advocate focusing on the psychological traits themselves rather than a concept of person defined by those traits (and as yet

unsatisfactorily defined). In response to Nelson's proposal, separating the two versions is easy enough to do but seems to accomplish little. What work would the concept of "person" be doing beyond being a new synonym for "normal adult human" (see also Blackford 2007)? It makes sense to talk about "persons" rather than "normal adult humans" only if some broader set of instances is encompassed, with at least some of the "hard cases." Indeed, much of the motivation for using the concept "person" comes from our sense that certain moral generalizations extend beyond the narrow set of normal adult humans.

One aspect of our position that may need additional clarification is the role of neuroscience in our argument. Although the evidence for the activity of a personhood network comes mainly from neuroscience, in principle neuroscience is no more relevant than psychology, from which we also drew support in our paper. We are not making any claims regarding the unique ability of neuroscience to adjudicate questions of personhood (Buford and Allhoff 2007; Racine 2007). But insofar as neuroscience does contribute informative data, we gratefully acknowledge the friendly amendments to our description of the person network (Phelps 2007).

The relation between empirical facts and moral theory was emphasized by several of the commentators, some of whom distinguished the metaphysical and empirical senses of personhood (Buford and Allhoff 2007; Nelson 2007; Meyers 2007; Racine 2007; Sagoff 2007). A variant of this point was the distinction between socially defined and empirical concepts (Banja 2007; Perring 2007). Of course, the whole program of naturalizing philosophical concepts, advocated by far better philosophers than we, involves collapsing or at least aligning such distinctions in the sense of finding the empirical correlates of metaphysical or moral concepts.

By now we hope it is clear that we are not advocating a naturalized conception of personhood. What we are doing is addressing the relation between the moral concept of a person and the natural world, and in that sense we are assuming that the natural world is relevant to moral theory. *Although moral principles themselves may not require empirical validation, they do refer to entities in the real world, and for bioethics in particular the way in which we anchor such principles in empirical reality is crucial.* One of Blackford's (2007) points about Lockean personhood is an example of a statement about the relation between the metaphysical realm of ethics and the real world: He pointed out that although Locke's definition of personhood is perfectly explicit and non-arbitrary, it is not sufficient to guide practical ethics because entities that are non-persons on Locke's view still deserve moral protection.

Given the dim prospects for clarifying the concept of a person, we suggest dispensing with that concept and instead basing our moral judgments on the interests of entities with varying levels of awareness, including self-awareness. Some commentators felt that this approach would end up encountering the same problems, because then we would need to determine whose interests should we care about (Glannon 2007; Grey, Hall and Carter 2007; Meghani 2007;

Perring 2007). As strange as it sounds to say that consciousness might be the easier of two things to assess, we believe that is. Compared to personhood, there is less doubt that it exists, and greater agreement as to the ranking of different entities. We take Fins' (2007) commentary to be consistent with this position. Rather than worry about whether a PVS patient is a person or not, we should focus on determining the level of awareness of such patients, many of whom languish in long-term care facilities with attention paid only to their physical needs.

Centuries of intellectual effort have gone into the explication of personhood. It is therefore quite a challenge to say something useful and new that might have escaped so many greater minds than ours. Our target article is no more than the beginning of an argument, suggested by scientific evidence not available to previous generations. We are grateful to the fifteen commentators for helping us to hone that argument by highlighting those aspects in need of clarification, revision, and further evidence.

#### ACKNOWLEDGMENTS

We thank Joshua D. Greene for helpful comments on an earlier draft of this Response.

#### REFERENCES

- Banja, J. 2007. Personhood: Elusive but not illusory. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 60–62.
- Blackford, R. 2007. Differing vulnerabilities: The moral significance of Lockean personhood. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 70–71.
- Buford, C., and F. Allhoff. 2007. Neuroscience and metaphysics (redux). *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 58–60.
- Churchland, P. S. 2007. The necessary-and-sufficient boondoggle. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 54–55.
- Dennett 1978.
- Farah, M. J., and A. S. Heberlein. 2007. Personhood and neuroscience: Naturalizing or nihilating? *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 37–48.
- Fins, J. J. 2007. Border zones of consciousness: Another immigration debate? *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 51–54.
- Glannon, W. 2007. Persons, metaphysics and ethics. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 68–69.
- Grey, W., W. Hall, and A. Carter. 2007. Persons and personification. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 57–58.
- Meghani, Z. 2007. Is personhood an illusion? *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 62–63.
- Meyers, C. 2007. Personhood: Empirical thing or rational concept? *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 63–65.
- Nelson, J. L. 2007. Illusions about persons. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 65–66.
- Perring, C. 2007. Against scientism, for personhood. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 67–68.

Phelps, E. A. 2007. The neuroscience of a person network. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 49–50.

Racine, E. 2007. Identifying challenges and conditions for the use of neuroscience in bioethics. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 74–76.

Roskies, A. L. 2007. The illusion of personhood. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 55–57.

Saghoff, M. 2007. A transcendental argument for the concept of personhood in neuroscience. *American Journal of Bioethics (AJOB-Neuroscience)* 7(1): 72–73.