Transhumanism and Personal Identity

James J. Hughes

Enlightenment values are built around the presumption of an independent rational self, citizen, consumer and pursuer of self-interest. Even the authoritarian and communitarian variants of the Enlightenment presumed the existence of autonomous individuals, simply arguing for greater weight to be given to their collective interests. Since Hume, however, radical Enlightenment empiricists have called into question the existence of a discrete, persistent self. Today neuroscientific reductionism has contributed to the rejection of an essentialist model of personal identity. Contemporary transhumanism has yet to grapple with the radical consequences of the erosion of liberal individualism on their projects of individually chosen enhancement and longevity. Most transhumanists still reflect an essentialist idea of personal identity, even as they embrace projects of radical cognitive enhancement that would change every constituent element of consciousness. Transhumanists need to grapple with how their projects and ethics would change if personal identity is an arbitrary, malleable fiction.

Personal Identity and the Enlightenment

The Enlightenment thinkers attempted to move past the idea of human nature as being defined by God-given immortal souls inhabiting flesh, to the view that we are rational minds emerging out of and transforming nature. For John Locke, for instance, an immaterial and supernatural soul was not a necessary explanation for the self. He proposed that it was within God's power to create matter that could think, and that it is our capacity to think which makes us persons. By setting aside the immaterial soul as the basis of identity however there was a problem with the identity of a person at the Resurrection of souls at the Judgment. If consciousness resides in the flesh, and the reconstituted body at the end of things has none of the stuff of the original body, then how could it be the same person? Because the mind in that flesh would remember its previous self. For Locke memory connected one's present self to one's past, and was therefore the basis of personal identity.

> …to find wherein personal Identity consists, we must consider what Person stands for; which, I think, is a thinking intelligent Being, that has reason and reflection, and can consider itself as itself, the same thinking thing in different times and places… (Locke, 1689)

Remarkably Locke also considers the problem of the splitting of personal identity through the example that consciousness might reside in a severed finger, and suggests that both the body and the finger could then have its own personhood. But the limb that had a continuous subjective identity is what is crucial for "in this personal identity is founded all the right and justice of reward and punishment." Thinking that we are the same person over the time was essential for moral accountability. We have to believe that we are the same person who acted justly or wrongly in the past, and will be punished or rewarded in the future.

But the further investigation of the nature of memory and consciousness almost immediately began to erode this idea of personal identity. Fifty years after Locke, the Scottish philosopher David Hume's dissection the self takes the new materialism and empiricism in a more radical direction. In Hume's analysis our belief that anything had an enduring substance was a perceptual illusion. In his Treatise on Human Nature he applies this skepticism to personal identity, and argues that the self is also an illusion

created by the contiguity of sense perceptions and thoughts. The self is merely a "…a bundle or collection of different perceptions which succeed one another with an inconceivable rapidity and are in perpetual flux and movement" (Hume, 1739). While for Locke memory was the core of personal identity, knitting together past and present self, for Hume memory was what created the illusion that there was some kind continuity between past and present mental states.

Hume did not pursue the conclusion his view of the self might have for moral accountability or political theory, and for understandable reasons. Locke's materialist psychology proposed that we had real, substantial selves grounded in our memory, and that these selves had interests that we could know and pursue. But Hume rejection of personal identity was incompatible with the Enlightenment project of building a new society of rational individuals pursuing self interests through democracy and market exchange. If we are as confused about the very nature of our selves, as Hume proposed, it calls into question a society founded on the equality of citizens, morally accountable persons, and individual rights, and would instead validate benevolent despotism towards collective goods.

Humeian skepticism about personal identity also was, and remains, deeply anti-intuitive. Almost all human beings are fairly certain of their personal identity over time. The spread of liberal individualism along with market economies, liberal political regimes, and expanding realms of individual choice have only made the intuitive belief in the continuous self, whether rooted in an assumed supernatural substance or simply materialism, even stronger.

The contradiction between the Enlightenment's concept of Lockeian selfhood, foundational to liberal individualism, and the Humeian empiricist recognition that the self is a fiction lay dormant until the twentieth century when neuroscience, another product of the Enlightenment, revived the debate. As neuroscientist collected accounts of patients with lesions and degenerative diseases they began to create an empirical model of the ways that the brain creates the ongoing narrative of the self, and illustrate just how malleable and fragile that narrative is. Multiple personality disorders are simply an extreme of the fractured tumult of desires and self images in all brains, validating Hume's claim that our personalities are more like parliaments than monarchs. Patients with severed corpus callosums can have their left hemispheres pursue goals separate from and contradictory to those being pursued by the right hemispheres. As philosopher Thomas Metzinger has documented so brilliantly in his 2009 *The Ego Tunnel: The Science of the Mind and the Myth of the Self*, neuroscience shows that the "self-y" feeling is simply a useful heuristic that our minds create, without any underlying reality. (See also Noe, 2009).

Similarly behavioral economists have carried out a similar deconstruction of the idea of the rational, utility-maximizing individual, showing that we our preferences are not autonomous or coherent, and our behavior generally irrational. For instance Daniel Kahneman's work on the experiencing versus remembered self shows that our memories of our lives are fictional narratives that bear little relationship to our actual moment to moment experience (Redelmeier, Katz and Kahneman, 2003).

Moral and political theorists have been slow to respond to this erosion of the core assumptions of Western thought. One of the few exceptions is the Oxford philosopher Derek Parfit, whose landmark 1986 *Reasons and Persons* argued a Humeian account of personal identity, but connecting it to a utilitarian moral theory. For Parfit there are is no substantial self, only correlations between our mental states at any two times, and that correlation declines over time. The self exists only insofar as an entity like England exists – it has a physical history and on top of that an evolving set of cultural groups and political institutions. But any attempt to definitely say that England began at a particular time and constitutes a specific set of people and institutions would simply be an arbitrary fiction.

For Parfit the moral and political upshot of our declining relationship to all future versions of our selves is that we have a corresponding interest in the welfare of all other future persons. At some point our future

selves are likely to be so dissimilar from our current selves that we are better off acting in the interest of all 80 year olds instead of simply our own. Of course even critics who accepted a declining self-similarity over time have scoffed at the possibility that an individual would ever bear as much similarity to other future persons as to the future person in their body.

With that cue, the transhumanist project of cognitive and biological enhancement enters the equation and proposes not only many more years over which the ties of personal identity could attenuate, but radical changes to desire, memory, cognition and identity which will fundamentally challenge our presumptions about the self.

Enhancement, Transhumanism and Personal Identity

Transhumanism has inherited many ideological contradictions from Enlightenment philosophy (Hughes, 2010). These include the conflict between atheism and the belief that intelligence could become godlike, or between teleological techno-optimism and the rationalist acknowledgment that the future is uncertain. The conflict between the Lockean and Humean views of the self is another Enlightenment contradiction working away within transhumanism. But the personal identity conundrum is perhaps more exclusive to transhumanism than the other intra-Enlightenment debates since it is precisely the prospect of radical neuroscience that has made the erasure of the illusion of personal identity so tangible.

Oxford's transhumanist philosopher Nick Bostrom acknowledged the problem of personal identity for transhumanism in the 2003 Transhumanist FAQ:

> Many philosophers who have studied the problem think that at least under some conditions, an upload of your brain would be you. A widely accepted position is that you survive so long as certain information patterns are conserved, such as your memories, values, attitudes, and emotional dispositions, and so long as there is causal continuity so that earlier stages of yourself help determine later stages of yourself.... These problems are being intensely studied by contemporary analytic philosophers, and although some progress has been made, e.g. in Derek Parfit's work on personal identity, they have still not been resolved to general satisfaction. (Humanity+, 2003)

In her 2009 essay "Future Minds: Transhumanism, Cognitive Enhancement and the Nature of Persons" Susan Schneider cites transhumanist Ray Kurzweil's 2005 parsing of the personal identity debate into four positions:

> **1. The ego theory**—a person's nature is her soul or nonphysical mind, and this mind or soul can survive the death of the body.
> **2. The psychological continuity theory**—you are essentially your memories and ability to reflect on yourself (Locke) and, more generally, your overall psychological configuration, what Kurzweil referred to as your "pattern."
> **3. Materialism**—you are essentially the material that you are made out of—what Kurzweil referred to as "the ordered and chaotic collection of molecules that make up my body and brain"
> **4. The no self view**—there is no metaphysical category of person. The "I" is a grammatical fiction (Nietzsche). There are bundles of impressions but no underlying self (Hume). There is no survival because there is no person (Buddha, Parfit).
> (Schneider, 2009)

In *The Singularity is Near* Kurzweil (2005) advocates for position Two, which has been dubbed "Patternism," and this is the dominant view among transhumanists in general. Patternism permits radical changes to the body and brain so long as the sense of continuity, the memory of a flow of mental states leading to the present, are maintained. Even something as radical as the recording of a personality in a brain and its reinstantiation in a computer would count as personal identity if the mind in the computer remembered the process leading to the change and identified with the prior biological person.

One transhumanist philosopher who defends the variant of the patternist view in great depth is Max More the founder of the Extropian school of transhumanist thought. More wrote his doctoral thesis on Derek Parfit's personal identity arguments and their implications for radical human enhancement. Although largely arguing that his view is consistent with Parfit's anti-essentialism, in the end More (1995) argued that so long as the radically transformed person was consistent with, or a fulfillment of, the values of the prior person, then personal identity had been maintained. More specifically argues against a focus on continuity of memory as important for identity. For More's "transformationalist" account values are the core of identity for most of us. On the other hand

> This is not true for everyone…Some persons lack a strong core of values. These persons would give up their identity through transforming… Those who value self-transformation strongly can undergo more changes in other characteristics while maintaining identity. (More, 1995)

In other words the pattern that determines personal identity are strongly held values, especially the value of self-transformation, and those without these strong values are at great risk of losing personal identity as they undergo enhancement.

Transhumanist philosopher Mark Walker (2008) also grapples with the personal identity objection to cognitive enhancement and adopts an implicitly patternist view. Walker says that since radical changes might violate personal continuity the path from humanness to "godlike" posthumanity should be gradual.

> If we must accept gradualism then the worst consequence is that it seems to slow down the process whereby one might become a posthuman, it does not prohibit it. In terms of the neural surgery experiment, we might imagine that if too many neurons are added to your brain at once you will cease to exist, but if neurons are gradually added your identity will be preserved. Accepting this means that I could not demand as a right to be upgraded to a posthuman overnight, but I could consistently demand as "my right" the right to a number of small interventions that would eventually lead to me becoming a posthuman.
> … we may be able to autonomously determine our identity through the exercise of technology on our biology. In fact this would be a higher expression of our autonomy than we can achieve today. (Walker 2008)

Both More and Walker concede that some enhancements would break personal identity by breaking the continuity of the personality pattern. But both believe, unlike many bioconservative critics, that personal continuity is nonetheless possible.

Schneider (2009) suggests however that the transhumanists' patternist theories are inadequate to establish the continuity of personal identity after radical cognitive enhancements or uploading. Transhumanist enhancement scenarios propose radical malleability in memory, values and all other elements of the "pattern." Transhumanists also accept the plausibility, even inevitability, of multiple copies of personalities which would all feel identity with the prior original person. While most transhumanists don't see the multiplication of selves as problematic, it is usually considered incompatible with the assumed transitive unity of identity over time. If there can be more than one You, do you really exist in the first place. She concludes by asking

> …what is it that ultimately grounds your decision to enhance or not enhance if not that it will somehow improve who you are? Are you perhaps merely planning for the well-being of your closest continuent? (Schneider, 2009)

Schneider does not address the fact that there are transhumanists, such as myself, who have taken position that the self is an illusion (Hughes, 2011, 2005). I have argued that radical longevity and cognitive enhancement will push liberal democratic society to adopt post-liberal individualist moral, legal and political frameworks that do not assume personal identity, although the contours of what such frameworks might be I cannot say. It is hard to discern however what meaning "liberty, equality and fraternity" would have without the convenient fiction of autonomous individuals as citizens.

Parfit's no-self utilitarianism, in which only the interests of all future persons, and not one's own personal identity, are taken into account is one such possibility.   It would then be possible to argue that even though personal identity is a fiction, that the good of the collective future person would be improved by maintaining the fiction of personal identity in life and law. This is similar to the debate over free will and legal culpability in the light of neuroscience; even if neuroscience demolishes the idea that any criminal truly chooses a criminal act, social utility will be greater if we pretend that individuals have moral choice and are accountable for their actions.

Many other accomodations to the erosion of personal identity can be imagined however, from efforts to use to neurotechnologies to create and rigid secure personal invariability, to their use to replace individual identity with completely collective identity (e.g. "the Borg"). The erosion may come about without any coercion; the selective suppression of the brain mechanisms that create the illusion of the self, such as proprioception, will likely be attractive targets for people exploring neurotechnologies for therapeutic and recreational reasons. Experiences previously accessible only to yogis, such as body boundlessness, empathic unity with others, or absorptive concentration will likely become commonplace. The recording of memory and experience will also enable the sharing of memory and experience with others, which many will want as a means to closeness, evangelism or simply vanity. How much of someone else's life would one need to remember before it called into question one's own identity?   By experimentally modifying our values and desires we could become people we wouldn't have wanted to be previously, from amoral to supermoral.

Nick Bostrom seems to have this kind of development in mind in his 2001 essay "Existential Risks." After discussing natural and technological threats that could wipe humanity out he addresses the threat of "shrieks" and "whimpers," futures in which our descendents still exist but not in forms which have maintained some essential continuity with who we are today. In 2004 he proposed a shriek scenario that might result from voluntary use of enhancement in the pursuit of capitalist competition.

> We can thus imagine a technologically highly advanced society, containing many sorts of complex structures, some of which are much smarter and more intricate than anything that exists today, in which there would nevertheless be a complete absence of any type of being whose welfare has moral significance.   In a sense, this would be an uninhabited society.   All the kinds of being that we care even remotely about would have vanished… the catastrophe would be that such a world would not contain even the right kind of machines, i.e. ones that are conscious and whose welfare matters. (Bostrom, 2004)

In other words, if we adopt a personal identity theory in regards the human project then some forms of post-personal identity societies might also be societies that no longer represent the human project. But then, the critique of identity essentialism probably applies at the level of society even more clearly than

for individuals. If there is no real self and no real humanity then we are left with the question of whether we want to collectively pretend that we do exist, and if so, to what ends?

References

Bostrom, Nick. 2001. "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," Journal of Evolution and Technology 9(1).

… 2004. "The Future of Human Evolution," in *Death and Anti-Death: Two Hundred Years After Kant, Fifty Years After Turing*, ed. Charles Tandy (Ria University Press: Palo Alto , California , 2004): pp. 339-371.

Hughes, James. 2001. "The Future of Death: Cryonics and the Telos of Liberal Individualism," *Journal of Evolution and Technology* 6(1).
http://jetpress.org/volume6/death.html

… 2005. "The illusiveness of immortality," In C. Tandy (Ed.), *Death and anti-death, volume 3: Fifty years after Einstein, one hundred fifty years after Kierkegaard* (pp. XX–XX). New York: Ingram.

… 2010. "Contradictions From the Enlightenment Roots of Transhumanism," *Journal of Medicine and Philosophy* 35 (6):622-640.

Humanity+. 2003. Transhumanist FAQ.
http://humanityplus.org/learn/philosophy/faq

Hume, David. 1739. A Treatise of Human Nature: Book I: Of the understanding. Part IV: Of the sceptical and other systems of philosophy. Section VI: Of Personal Identity.
http://www.mnstate.edu/gracyk/courses/web%20publishing/TreatiseI.iv.vi.htm (accessed Dec 5, 2009)

Kurzweil, R. (2005). *The Singularity is Near: When Humans Transcend Biology*. New York: Viking.

Locke, John. 1689. Essay Concerning Human Understanding.
http://oregonstate.edu/instruct/phl302/texts/locke/locke1/Essay_contents.html

Metzinger, Thomas. 2009. *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. Basic Books.

More, Max. 1995. "The Diachronic Self. Identity, Continuity, Transformation. PhD dissertation, University of Southern California.
http://www.maxmore.com/disscont.htm

Noe, Alva. 2009. *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. Hill and Wang.

Parfit, Derek. 1986. *Reasons and Persons*. Oxford University Press.

Redelmeier D.A., Katz, J., & Kahneman, D. 2003. Memories of colonoscopy: A randomized trial. *Pain*, 104: 187-194.

Schneider, Susan. 2009. "Future Minds: Transhumanism, Cognitive Enhancement and the Nature of Persons," In: *The Penn Center Guide to Bioethics* (pp. 844-856), Vardit Ravitsky, Autumn Fiester and Arthur L. Caplan (eds.). Springer.

Walker, Mark. 2008. " Cognitive Enhancement and the Identity Objection," *Journal of Evolution and Technology* 18(1): 108-115.
http://jetpress.org/v18/walker.htm