

Sorgner wrote “significant similarities between the posthuman and the overhuman can be found on a fundamental level” (Sorgner 2009, 1), in reaction to Bostrom’s claim that there are only “surface-level similarities” between these concepts (Bostrom 2005, 4). While there certainly must be some similarity between the posthuman and the overhuman, there is a fundamental difference that posthumans will be a reality, whereas the overhuman is an ideal limit that can never be reached.

### **Nietzsche’s overhuman**

Nietzsche described the overhuman as follows:

Here man has been overcome at every moment; the concept of the “overman” has here become the greatest reality – whatever was so far considered great in man lies beneath him at an infinite distance. (Nietzsche 1888, 305.)

The important point is the “infinite distance” between human and overhuman. As envisioned by the transhumanist community, posthumans are a finite improvement on humans.

Nietzsche's overhuman is closely related to his concept of "eternal recurrence." Faced with the prospect of living one's life again endlessly, with every detail and misery replicated exactly, the ordinary human says no but the overhuman says yes. Nietzsche believed in human improvement, driven by a human "will to power." But the overhuman has no need for improvement, having achieved satisfaction with life. The overhuman is an ideal rather than an achievable reality. Posthumans, as envisioned by most transhumanists, will be real successors to humans and still struggling to improve.

### Analogy with fixed point theory

The notion of human and posthuman improvement suggests a partial order relation among sentient beings. The notion of beings wanting to improve suggests a function from beings to the beings they want to become. And the notion of the overhuman being satisfied with eternal recurrence of its life suggests a fixed point of this function. Taken together, these suggest an analogy with the mathematical theory of fixed points over partially ordered sets, as used in the denotational semantics of programming languages (Scott 1972). We should not take this analogy too seriously because of the lack of rigorous definitions of posthuman, overhuman and other terms, but it is nevertheless interesting.

Let  $S$  represent the set of possible sentient beings, including humans, posthumans and the overhuman. We define a partial order relation on  $S$ , writing  $s < s'$  for  $s$  and  $s'$  in  $S$ , to mean that  $s'$  is an improvement on  $s$ . This is a partial order because it may be the case that neither  $s < s'$  nor  $s' < s$  is true (e.g., neither my wife nor I am an improvement on the other, in the interest of domestic tranquility). We also define a function  $F: S \rightarrow S$  where  $F(s)$  is the being that  $s$  wants to become. Humans want to improve, so for all human  $s$ ,  $s < F(s)$ . Similarly a posthuman  $s$  as envisioned by the transhumanist community will want to improve, so for these posthumans  $s < F(s)$ . This assumes that all humans and posthumans agree on what constitutes improvement, which may not be true. Nietzsche discussed how different people had different moralities but he imposed on this discussion his own standard of strength and weakness among humans, which was his real definition of improvement. And there is a general consensus among humans on at least certain types of improvement, so perhaps we can restrict  $S$  to those sentient beings who subscribe to this consensus. An objective mathematical measure of intelligence has been proposed and might contribute to a definition of improvement (Legg and Hutter 2007).

The sentient being called the overhuman would happily relive its life in eternal recurrence, so overhuman =  $F(\text{overhuman})$ . That is, overhuman is a fixed point of  $F$ . Under certain assumptions about  $S$ , the order relation and  $F$ , the Kleene fixed point theorem tells us that a least fixed point of  $F$  exists and is defined as the limit of the infinite sequence:  $\perp < (\perp \leq F(F(\perp)) \leq F(F(F(\perp))) \leq \text{etc.}$  (Manes and Arbib 1986). Here  $\perp$  indicates the least member of  $S$ , whose existence is one of the theorem's assumptions. Since we are looking for a fixed point greater than humans (i.e., the overhuman) rather than a least fixed point, we can replace  $\perp$  by any human  $s_1$  (if we restrict  $S$  to those  $s$  such that  $s_1 \leq s$ , then  $s_1$  is the least member).

Another assumption of Kleene's theorem is that  $S$  is a complete partial order, which means that every directed subset (every pair of members of a directed subset has an upper bound in the subset) of  $S$  has a least upper bound in  $S$ . Since  $S$  is a set of possible rather than actual beings, it is plausible that there exists a possible being that is a least upper bound for every directed subset of  $S$ .

The final assumption of Kleene's theorem is that  $F$  is continuous, which means that it preserves least upper bounds. That is, given a directed subset  $D$  of sentient beings then  $F(D)$  is the set of beings that each being in  $D$  wants to become. We can also define  $d$  and  $d'$  as the least upper bounds of  $D$  and  $F(D)$ . Continuity of  $F$  means that  $d'$  is the being that  $d$  wants to become (i.e.,  $d' = F(d)$ ). If  $F$  satisfies this assumption, and if  $S$  is complete, then Kleene's theorem says we can define the overhuman as the limit of the infinite sequence of improving humans and posthumans:  $s_1 \leq s_2 = F(s_1) \leq s_3 = F(s_2) \leq s_4 = F(s_3) \leq \text{etc.}$

We are assuming that all humans and posthumans want to improve, so this sequence should be strictly increasing:  $s_1 < s_2 = F(s_1) < s_3 = F(s_2) < s_4 = F(s_3) < \text{etc.}$  Thus the analogy with fixed point theory suggests that the overhuman's satisfaction with eternal recurrence implies that the overhuman must be the result of

an infinite sequence of improving humans and posthumans. This is consistent with the Nietzsche quote that humans are infinitely far beneath the overhuman.

However, according to current physics the universe has a finite information capacity (Lloyd 2002), so there cannot be an infinite sequence of strict improvement. Some  $s_n$  in the sequence of improving posthumans must reach the maximal improvement possible in our finite universe. Assuming that this  $s_n$  is intelligent enough to realize that it has reached the maximum, it will not want to improve (violating our assumption that all posthumans want to improve), so  $s_{n+1} = F(s_n) = s_n$ . That is,  $s_n$  is a fixed point of  $F$ . If current physics is right that our universe is finite, then it is possible that some posthuman would say yes to eternal recurrence. But nineteenth century physics did not derive any finite limit on the information capacity of the universe and a finite limit on the overhuman is inconsistent with the Nietzsche quote.

The posthuman  $s_n$  that reaches maximal improvement is reminiscent of Tipler's prediction of human intelligence spreading to employ all the matter and energy in the universe (Tipler 1994). But the goal of Tipler's work was to derive Christian theology from physics, so Nietzsche would probably have denied any connection with his overhuman (perhaps Newton would have approved of Tipler's goal).

This analogy with fixed point theory is one interpretation of Nietzsche's overhuman and eternal recurrence. His romantic writing style is open to many interpretations.

### Radical inequality

Sorgner writes about social issues including the dangers of genetic engineering, people's concern for their children, and eugenics (Sorgner 2009). But he does not address the issue of the radical inequality that could result from technological change to human bodies and brains. Some transhumanists think that this is a critical issue (Hughes 2004; Hibbard 2008a), whereas others focus only on individual improvements without regard for the effects of unequal access to the technologies of improvement. When humans can simply buy greater intelligence and use that intelligence to earn more money, this positive feedback cycle will lead to an unstable "arms race" in intelligence (Hibbard 2008b). Intelligence levels among humans will diverge to the extent that less intelligent humans will be unable to understand or learn the languages spoken by the most intelligent humans, leading to different laws for people of different intelligence. This must have a destructive effect on the sense of meaning in the lives of less intelligent people.

Nietzsche thought that strength was the ultimate good and expressed little sympathy for measures to oblige the strong to subsidize the weak. Thus Nietzsche is not a good antecedent for transhumanists concerned with the issue of radical inequality. Hobbes is a more useful antecedent than Nietzsche for such transhumanists. Hobbes was a materialist with a practical writing style who wrote "By reasoning, however, I understand computation" (Hobbes 1981, 177). So Hobbes would probably not be surprised that 350 years later humans are approaching the construction of machines surpassing the human mind. More important for transhumanism is Hobbes's observation that humans need stability and security, but that society will degenerate into chaos without a social contract and an authority to enforce that contract (Hobbes 1968). The technology of mind will invalidate assumptions underlying our society and lead to instability and insecurity unless we modify our social contract to regulate that technology (Hibbard 2008a). Following Hobbes, transhumanists should analyze this situation and ask what social contract will create stability and security for people to live meaningful lives.

### References

- Bostrom, N. 2005. A history of transhumanist thought. *Journal of Evolution and Technology* 14(1): 1-25.
- Hibbard, B. 2008a. The technology of mind and a new social contract. *Journal of Evolution and Technology* 17(1): 13-22.

- Hibbard, B. 2008b. Adversarial sequence prediction. In *Artificial general intelligence 2008, Proceedings of the first AGI conference*, ed. P. Wang, B. Goertzel and S. Franklin, 399-403. Amsterdam: IOS Press.
- Hobbes, T. 1968. *Leviathan*. Baltimore: Penguin Books, 1968. (First published 1651.)
- Hobbes, T. 1981. *Part I of De corpore*. Trans. A. Martinich. New York: Abaris, 1981. (First published 1655.)
- Hughes, J. 2004. *Citizen cyborg: why democratic societies must respond to the redesigned human of the future*. Cambridge, MA: Westview Press.
- Legg, S. and M. Hutter. 2007. Universal intelligence: A definition of machine intelligence. *Minds and Machines* 19(4): 391-444.
- Lloyd, S. 2002. Computational Capacity of the Universe. *Physical Review Letters*. 88, 237901.
- Manes, E., and M. Arbib. 1986. *Algebraic approaches to program semantics*. New York: Springer.
- Nietzsche, F. 1989 1888. Ecce homo, how one becomes what one is. Trans. Walter Kaufmann, in On the genealogy of morals and Ecce homo, 201-335. New York: Vintage Books. 1989. (First published 1888.)*
- Scott, D. 1972. Lattice theory, data types and semantics. In *Formal semantics of programming languages*, ed. Randall Rustin, 65-106. Englewood Cliffs, NJ: Prentice-Hall.
- Sorgner, S. L. 2009. Nietzsche, the overhuman, and transhumanism. *Journal of Evolution and Technology* 20(1): 29-42.
- Tipler, F. 1994. *The physics of immortality*. New York: Doubleday.

DR. RUPNATHJI (DR. RUPAKNATHJI)