Response to Commentary

Response to the Commentary of Nils J. Nilsson on "Human Consciousness and Selfhood"

David Sahner[*]

ABSTRACT

This is my Response to Nils J. Nilsson's Commentary on my essay "Human Consciousness and Selfhood: Potential Underpinnings and Compatibility with Artificial Complex Systems" which appeared in the December 2010 issue of JCER.

Key Words: human consciousness, selfhood, artificial intelligence, complex system.

Nils Nilsson has my gratitude for his perceptive comments. In truth, our positions are not as far apart as one might initially think. I completely agree that human consciousness comes in many flavors. For the most part, I would consider these variations to constitute separate "subspecies" of a given type of consciousness. Given (a) the significant differences between the substrates within which AI and human intelligence are, respectively, embedded, and (b) the virtually infinite complexity of the (importantly perishable) human central nervous system and body, I strongly suspect, however, that machine consciousness will be at much more of a remove from human consciousness than, say, my consciousness is from that of a native of Japan. Perhaps it may be best to think of machine consciousness, if and when it should arise, as belonging to a different genus or phylum. Admittedly, this is largely opinion because the riddle of human consciousness has not been deciphered to everyone's satisfaction. Where, precisely, robot consciousness will fit into the taxonomy of consciousness must be empirically defined in the future.

The point is made by Nils Nilsson that computers can glean knowledge in ways that have nothing to do with language, and that this does away with a barrier I cite to the instantiation of human intelligence, and attendant consciousness, in an artificial complex system. I would not dispute the statement that machines are capable of various non-linguistic forms of learning, and I am familiar with the fantastic vistas that have been opened by neural nets. What I would emphasize here is that knowledge of *specifically human experience* cannot be faithfully imparted to a computer because, for the reasons cited in my first paragraph above, it will be of a fundamentally different nature – and language, even the language of the greatest poets, provides only the shadow of a unique human experience. I have no doubt that machines can acquire "knowledge" in many ways, in some cases far more efficiently than humans. In the future, machines may have "machine knowledge" of which they are consciously aware (and which may be terrifically useful) but I don't believe machine consciousness, if it comes to pass, will be precisely like mine. Software may now exist that can reverse engineer a Bach-like piece based on stylistic patterns evident in his work, but I'd claim that this is "discrimination" - not necessarily "sensation." The computer program may be able to abstract a "Bach-like" signature technique from its high-level analysis of a series of notes, but that perception does not necessarily entail sensation as defined by Nicholas Humphrey.

Correspondence: David Sahner, M.D., Aeneas Medical Consulting, LLC. E-mail: davidsahner@yahoo.com

Most importantly, I think the field of AI will have to grapple with the virtually infinite, chaotic, and changeable metabolic and organizational complexity that underlies human sensation - and this does not reside merely in the brain, of course. This is extensively covered in section D in the list of challenges faced by AI outlined in my manuscript. For the reasons stated in my paper, I don't think computational approaches will be able to faithfully mimic the goings on in the human nervous system. Fold in the remainder of the human body serving as the link between a mutable environment (and its uniquely human cultural trappings) and the central nervous system, and the cliff to climb becomes even more glassy and vertically inclined. In truth, the interactions are bidirectional among components of this triad (i.e., body with mind, mind with environment through the body). Human meaning is tied to metaphors that are grounded in uniquely human physicality, notwithstanding the existence of machines with, for example, more sensitive auditory sensors. In the context of a highly stimulating string of email exchanges with Nils Nilsson, he assessed the "complexity" argument adduced above (and far more explicitly described in my manuscript) as the strongest one on offer in my paper.

In summary, I would not deny the possibility of machine consciousness in the future. My contention is more modest. I don't believe that consciousness, if it does evolve in machines, is at all likely to faithfully replicate the human brand of consciousness. That is to say, I have grave doubts about the likelihood of achieving broadly instantiated human-level AI (perhaps a much better term might be *"human-quality"* AI) and attendant *human* consciousness in an artificial complex system. In an effort to produce consciousness *in silico*, perhaps we can rely upon the precedent set by the evolution of human consciousness. That is, create a large number of "embodied" and self-replicating artificial complex systems with genuine and perishable "skin in the game." Note that I use that idiomatic expression with great purpose here. Engender competition for mates and resources and engineer a milieu in which cooperation among conspecifics is evolutionarily adaptive. Perhaps let these creatures interact for a few millennia and then, maybe one day, come back to ask one of them how it "feels" to be alive. You may get a very meaningful answer.

References

Nilsson, N. J. (2011), Commentary on David Sahner's "Human Consciousness and Selfhood." JCER V2(2): pp. 214-215.