

Consciousness and AI: A Reconsideration of Shanon

Tracy B. Henley

Mississippi State University

Shanon (1990) provides us with a well reasoned and careful consideration of the nature of consciousness. Shanon argues from this understanding of consciousness that machines could not be conscious. A reconsideration of Shanon's discussion of consciousness is undertaken to determine what it is that computers are missing so as to prevent them from being conscious. The conclusion is that under scrutiny it is hard to establish a priori that machines could not be conscious.

On the opening page of his recent essay "Consciousness" (Shanon, 1990) Professor Shanon remarks "The fact that human beings are conscious whereas computers are not underlines the significance of the phenomenon even further" (p. 137). In light of his own subsequent analysis I find his remark troubling. I would like to briefly review Shanon's argument to tease out more fully its implications for artificial intelligence computers (henceforth AI). I will then conclude with a briefer still explication of why this troubles me.

Shanon starts with the question "What is consciousness?" He explains that there are three aspects of consciousness. The first he calls *sensed being* or the ". . . difference between the being of organisms and mere thinghood of objects. Inasmuch as this difference is sensed, the organism is conscious" (p. 139). Without at all intending to be difficult I just am not sure what the difference between myself and my computer is supposed to be. What is it that I am supposed to sense and how would I know if I had failed to sense it? Could I be mistaken? And what possibly beyond observable responses would suggest that someone (or something) else had this sense? If it is just a matter of observable responses then surely a computer has just as much a claim to this sense as a person — that being — under certain (environmental) conditions they exist and we observe appropriate responses and under other conditions they do not exist or respond appropriately.

The second aspect of consciousness is *mental awareness*, which Shanon claims is characteristically human. This aspect of consciousness Shanon suggests is the ordinary usage of the term. However, as Shanon is aware, the ordinary use and philosophical extension of this term have two very different senses. One view, that of Locke, James, and anyone else who thinks seriously about it for a moment, is that consciousness is *that* we are directed to, perceiving, and doing. The content of consciousness then is frequently some subset of *what* we are directed to, perceiving, or doing. The other view holds that consciousness is the ability to report upon what we are directed to, perceiving, or doing. Beyond the fact that this later view creates all types of conceptual confusions that careful scholars (e.g., James, 1890; Wittgenstein, 1953) have wanted to avoid, it is just false. I am perfectly able to boot-up my computer, operate vending machines, walk through my world – indeed *drive* through my world – with a great deal of automaticity. Which is to say without any ability to report on the specifics of what I was directed to, perceiving, or doing, that guided all my actions. I can of course create *ad hoc* rationalizations of what I must have been seeing or doing given the outcome, but surely consciousness is something other than a set of *ad hoc* rationalizations.

The third aspect of consciousness is *reflection*, specifically in Shanon's own words, reflection "in the sense of a function taking its own value as an argument" (p. 140). Unless Shanon is suggesting something seriously mathematical here, and no argument of that type is provided, computers do this all the time. Most screen pauses that allow the user time to read a printed message, alarm clock functions, and a host of other common subroutines perform a recursive, self-monitoring functioning by constantly modifying their own argument. Even if what Shanon has in mind is the type of mathematics that Russell discounts, it should be noted that the computer has gone a long way toward establishing the computability of such functions.

Having defined consciousness Shanon proceeds to a consideration of the phenomenology of consciousness. First, I should note that Shanon and I seem to have a very different sense of the term "phenomenology." My sense of the term associates it with a *methodology* for conducting scientific or philosophical inquiry. In this sense James (1890) and Wittgenstein (1953) are providing phenomenologies of consciousness just as others (e.g., Dapkus, 1985; Thompson, Locander, and Pollio, 1989) have provided phenomenological accounts of other lived events. In this methodological sense neither Shanon or Hegel is very phenomenological.

Nevertheless, what Shanon has to say under the heading of phenomenology has fascinating possibilities. Among these possibilities for Shanon is that "Clearly, consciousness is a mental phenomenon" (p. 145). Shanon notes that Whitehead did not hold this view; but neither did John Watson, Clark Hull, Ludwig Wittgenstein, or B.F. Skinner. As such, it does not strike me either

then that consciousness is so *clearly* a mental phenomenon. Just stating something strongly does not make it so.

Nearer to the close of the essay Shanon makes an excellent observation that, with but a few exceptions (e.g., Heidegger, James, Vygotsky, Wittgenstein), has been omitted from pre-Gibsonian discussions of consciousness. That important point is the view that action-in-the-world is somehow basic to consciousness. I share Shanon's intuitions here and am persuaded by the arguments to which he refers (e.g., Gibson, 1979; Sudnow, 1980; Winograd and Flores, 1986). While Shanon does not move from placing a primacy upon "being-in-the-world" to a claim that machines, lacking bodies, are not conscious – because it is very clear that if they are not also in-the-world, I do not know where they are, or how I interact with them – he does make a parallel move that is equally vexing. Shanon invokes the most dubious and abused of all theoretical constructs, the self, as that which is to bridge the mental (conscious) and the physical (bodily). What is worse is that without any argument he states that "Any computational boxing of the notion of the self would certainly miss precisely what is special about conscious self" (p. 149).

My point in all of this is simple. While I have taken Professor Shanon to task here there is much to recommend his consideration of consciousness. There is nothing argued convincingly however to rule out consciousness in computers. I do want to be perfectly clear about my sincere respect of Shanon's ability to think comprehensively about a topic so diverse and difficult as consciousness. Indeed, beyond its overly rational approach I have little to quarrel with about Shanon's depiction of consciousness – only with his implications that it is something computers could not possess.

Others, such as Searle (1980), offer similar arguments with the clear intention of establishing a priori both that computers are not conscious and as such that any strong AI is impossible. If machines will ever be constructed that converse in natural language or that resemble the type of consciousness-seeming androids seen on television and film (e.g., Star Trek, Blade Runner) is however an empirical question that cannot be answered definitively in the here and now. Indeed, my intuitions about the relationship between consciousness and AI run just the opposite of Shanon's, namely, that the growing use of AI technologies in industry and the military will give rise to evermore sophisticated computers and that such computers will perform tasks that force us to ask new and more creative questions about the nature of consciousness.

Until then *en passant* remarks about what computers cannot do come across as little more than chauvanism. Perhaps it is human nature to find a certain solace in believing that we have some special quality that separates us from other entities such as computers (consider also the recent debates over animal

rights and animal consciousness). Such chauvanism amounts at best to bad epistemology. In the past we have overcome the bad epistemologies that have filtered down to influence our stance toward other things (such as persons of another sex or color) about whom we have convinced ourselves a priori that we are different from and superior to. As computers come to do more and more impressive things that cause us to continue to question the very nature of consciousness, let us hope we have learned from our past mistakes not to be premature about saying what is and is not like us without good empirical reasons to do so.

References

- Dapkus, M. (1985). A thematic analysis of the experience of time. *Journal of Personality and Social Psychology*, 49, 101-110.
- Gibson, J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.
- James, W. (1890). *The principles of psychology* (2 volumes). New York: Holt.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 417-457.
- Shanon, B. (1990). Consciousness. *The Journal of Mind and Behavior*, 11, 137-151.
- Sudnow, D. (1980). *Talk's body*. Hammondsworth: Penguin.
- Thompson, C., Locander, W., and Pollio, H. (1989). Putting the consumer back into consumer research. *Journal of Consumer Research*, 16, 133-146.
- Winograd, T., and Flores, C. (1986). *Understanding computers and cognition*. Norwood, New Jersey: Ablex.
- Wittgenstein, L. (1953). *Philosophical investigations*. Oxford, England: Basil Blackwell.