The Journal of Mind and Behavior
Summer 1991, Volume 12, Number 3
Pages 371-376
ISSN 0271-0137

Consciousness and the Computer: A Reply to Henley

Benny Shanon

The Hebrew University of Jerusalem

This paper is a response to Henley who criticizes a previous paper of mine arguing against my claim that computers are devoid of consciousness. While the claim regarding computers and consciousness was not the main theme of my original paper, I do, indeed, subscribe to it. Here, I review the main characteristics of human consciousness presented in the earlier paper and argue that computers cannot exhibit them. Any ascription of these characteristics to computers is superficial and misleading in that it fails to capture essential, intrinsic features of human cognition. More generally, psychological theory couched in terms of semantic representations and the computational operations associated with them is bound to be inadequate. The phenomenology of consciousness is a specific case marking this inadequacy.

My paper "Consciousness" (Shanon, 1990a; henceforth, CON) opens with the statement that whereas human beings are conscious, computers are not. Whether computers are endowed with consciousness, however, is not a question that CON intended to examine. This question bears on some of the most basic conceptual issues pertaining to the foundations of cognitive science, and undoubtedly it deserves serious, independent discussion. Indeed, the question has received extensive treatment in both cognitive and philosophical literature. The seminal paper in this domain of inquiry is Turing (1950), in which the question "Can computers think?" was posed. Of the many subsequent treatments of this, and related questions, one might single out Anderson (1964), Dreyfus (1979), Haugeland (1978) and Dennett (1979); none of these, I might note, is cited by Henley (1991).

What CON did set itself to do was to define the characteristic features of a given experience, namely, human consciousness. Specifically, CON attempted to describe a particular phenomenological domain, to specify the basic structural patterns it manifests, and to mark the internal structures it

I thank Ammon Levav and Gabi Trainin for stimulating discussion and for thier helpful comments. Requests for reprints should be sent to Benny Shannon, Ph.D., Department of Psychology, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel.

exhibits. It is in this sense that the examination pursued in CON is phenomenological. This sense of the epithet is different from the classical sense introduced by Husserl, but it is also different from Hegel's; for a general theoretical discussion and for a specific application the reader is referred to Shanon (1990b) and to Shanon (1989a), respectively.

The foregoing statement of interest and intent also defines the line to be taken in the present, brief note. While I do maintain that computers are not conscious, here I purport to present neither a full-fledged defence of this stance nor even a serious analysis of it. Rather, I would like to single out several key patterns manifested by human consciousness and to emphasize the difference between them and patterns exhibited by some computational systems of symbol information processing. The patterns to be singled out pertain to the three facets of human consciousness corresponding to the three types of consciousness introduced in CON and discussed by Henley: sensed being in the world, mental awareness and reflection.

Sensed being in the world. Is the computer in the world? Is it in touch with the world? *Prima facie*, it is. The computer may have a device that records information from the environment; it may also have the ability to exert control on the world and manipulate objects in it. Whether such interaction results in sensation is, as Henley points out, something no one can empirically determine. What I would like to point out, however, is that the human interaction with the world is categorically different from that which may be ascribed to a computer in the manner indicated above. First, for the computer, the tie with the world – be it efferent or afferent – is extrinsic. The transducers that relate the computer to the world outside are independent of the system that processes information. Except for the initiation of processing and its termination the computer operates without any interaction with the world. Second, the computer processes information by way of manipulating symbols. Whether these symbols have meaning or not, whether they relate to the world or not is, as far as the functioning of the computer is concerned, totally immaterial. The computer, in other words, lacks what is perhaps the key feature of the mind's tie with the world, namely, intentionality. Third, in line with observations made by both Gibson (1966, 1979) and his followers in the school of ecological psychology (see, for instance, Turvey and Shaw, 1979; Turvey, Shaw, Reed, and Mace, 1981) and by the biologically-oriented paradigm of autopoiesis (Edelman, 1987; Maturana, 1978; Maturana and Varela, 1980), it seems to me that the very definition of the mind, its structures and its modes of operation is intimately tied with the world. Human cognition, like biological organisms in general, is autopoietic: it constitutes its environment, and the environment constitutes it. The computer – at least in its present realizations – does not exhibit such dynamic interaction with the environment.

Mental awareness. Again, *prima facie* the construction of a computer endowed with awareness seems to be straightforward. Specifically, one could "colour" some information the computer entertains or some processing that it executes and mark them as being special. The question is whether such a marking affects in any way the manner in which the computer functions. I think not. By contrast, it seems to be that some aspects of human cognition are dependent on human beings having mental awareness. This is not the place to review functional benefits of mental awareness; for discussion the interested reader is referred to Shanon (1989b) and to Marcel and Bisiach (1988).

Reflection. Again, *prima facie* there is nothing special about reflection. As noted in CON, one could simply incorporate within a system's data-base information about the system and its current states and/or modes of operation. While such suggestions have been made in the literature (see Minsky, 1968), the "self" defined in this manner is very different from the human self. By way of illustrating this difference consider a device consisting of an information processing system coupled with a closed circuit television: the system views itself as it appears on the television screen. Such viewing, however, is not different from one's viewing of another person, or of any entity in the world, for that matter. Human self-reflection is categorically different. Experientially, our view of ourselves is unlike our viewing of any other person. On the one hand, we do not see ourselves; on the other hand, we know ourselves in a direct manner that no one else can. As suggested in CON, this knowledge brings together aspects that standard analyses (including computer-oriented analyses) would characterize as contradictory: the subject and the object, the bodily and the mental, the static and the dynamic.

In the foregoing discussion the three facets of consciousness were presented as three distinct types. As pointed out in CON, however, while qualitatively distinct, the three types are interrelated and together they form one unified whole manifesting coherent internal structures. Furthermore, human consciousness continuously vacillates between the three types, thus exhibiting what I have referred to as *resonance.* Any fragmented computer simulation of any of the three facets that does not capture the coherent dynamic structure that these define in unison cannot be deemed an adequate model of human consciousness or of the likes of it.

To all this, one might retort by saying that nothing in the foregoing comments implies that computers cannot exhibit consciousness. A sufficiently complex computer, construed in a dynamic, interactive fashion might eventually manifest the very phenomenology that has been sketched in CON. Whether such a computer will ever be built or not is a question on which I would not wish to speculate. As far as I am concerned what is important is to appreciate the characteristic patterns that such a system should manifest.

In my paper, I have presented indications that the system in question could not be one consisting of the manipulation of well-defined, well-formed symbols. Further, in the system in question there should be no segregation between data structures and the computational operations that apply on them, between information and the medium in which it is articulated, between subject and object, between cognition and its material realization. The present-day digital computer does impose such segregations; this is also true of the standard representational model of human cognition. Whether an alternative computer could ever be devised should be the concern of the designers of future-day technologies. I am a psychologist, and a conscious human being, and what I have attempted to do is present potential inventors with the specifications of what their products should meet if they are ever to be endowed with consciousness.

But is this not chauvinistic? Perhaps (as Henley insinuates) it is. Yet, it is a chauvinism founded in humility. The humility is two-fold. On the one hand, it consists of the avowal of ignorance: human consciousness is the only kind of consciousness we know and it is the only one we can discuss. On the other hand, the humility is an expression of awe and wonder: even though it is directly experienced by us all, human consciousness defies seemingly established dichotomies and categorizations standardly endorsed by both common sense and science. The drawing of hasty similarities between computer and human beings may be one of the many expressions of *Homo technologicus'* arrogant vanity that all of us, members of the species Homo sapiens, should endeavor to avoid.

References

Anderson, A.R. (Ed.). (1964). *Minds and machines.* Englewood Cliffs, New Jersey: Prentice Hall.

Dennett, D. (1979). Intentional systems. In D. Dennett, *Brainstorms* (pp. 3–22). Hassocks, Sussex: Harvester Press.

Dreyfus, H.L. (1979). *What computers can't do: A critique of artificial reason* (second revised edition). New York: Harper and Row.

Edelman, G.M. (1987). *Neural Darwinism.* New York: Basic Books.

Gibson, J.J. (1966). *The senses considered as perceptual systems.* Boston: Houghton-Mifflin.

Gibson, J.J. (1979). *The ecological aproach to visual perception.* Boston: Houghton-Mifflin.

Haugeland, J. (1978). The nature and plausability of cognitivism. *The Behavioral and Brain Sciences, 1,* 215–260.

Henley, T.B. (1991). Consciousness and AI: A reconsideration of Shanon. *Journal of Mind and Behavior, 12,* 367–370.

Marcel, A.J., and Bisiach, E. (Eds.). (1988). *Consciousness and contemporary science.* Oxford: Clarendon Press.

Maturana, H.R. (1978). Biology of language: The epistemology of reality. In G.A. Miller and E. Lenneberg (Eds.), *Psychology and biology of language and thought* (pp. 27–63). New York: Academic Press.

Maturana, H.R., and Varela, F.J. (1980). *Autopoiesis and cognition.* Dordrecht, Holland: Dreidel.

Minsky, M. (1968). Matter, mind and models. In M. Minsky (Ed.), *Semantic information processing* (pp. 425–432). Cambridge: MIT Press.

Shanon, B. (1989a). Thought sequences. *The European Journal of Cognitive Psychology, 1*, 129–159.
Shanon, B. (1989b). Why do we (sometimes) think in words? In K.J. Gilhooly, M. Keane, R. Logie, and G. Erdos (Eds.), *Lines of thought: Reflections in the psychology of thinking* (pp. 5–14). Chichester: John Wiley and Sons.
Shanon, B. (1990a). Consciousness. *Journal of Mind and Behavior, 2*, 137–152.
Shanon, B. (1990b). Non-representational frameworks for psychology: A typology. *The European Journal of Cognitive Psychology, 2*, 1–22.
Turing, A.M. (1950). Computing machines and intelligence. *Mind, 59*, 433–460.
Turvey, M.T., and Shaw, R. (1979). The primacy of perceiving: An ecological reformulation of perception for understanding memory. In L.G. Nillson (Ed.), *Perspectives on memory research: Essays in honor of Uppasla University's 500th anniversary* (pp. 167–222). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
Turvey, M.T., Shaw, R.E., Reed, E.S., and Mace, W.M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn. *Cognition, 3*, 237–304.