# Is consciousness necessary to high-level control systems?

## Maurizio Tirassa

Centro di Scienza Cognitiva • Università di Torino
via Lagrange, 3 • 10123 Torino (Italy)
tirassa@psych.unito.it

Istituto di Psicologia della Facoltà Medica, Università di Milano
via Francesco Sforza, 23 - 20122 Milano (Italy)

**Abstract.** Building on Bringsjord's (1992, 1994) and Searle's (1992) work, I take it for granted that computational systems cannot be conscious. In order to discuss the possibility that they might be able to pass refined versions of the Turing Test, I consider three possible relationships between consciousness and control systems in human-level adaptive agents.

## Introduction

Artificial Intelligence (AI) can be viewed as the attempt to build a person, under the specification that it (the resulting person) should be the necessary, predictable result of the carrying out of a set of computations. While there is no constraint on the nature of such computations, artificial brains cultured in nutrient media should be excluded from this definition.

Newell (1990) takes the mind to be "the control system that guides the behaving organism in its complex interactions with the dynamic real world" (Newell 1990, p. 43). This definition is intended for artificial as well as natural minds. The aim of AI, thus, is to devise computational versions of such control systems, together with suitable sensors, effectors, etc. Let us dub "robots" the products of this enterprise.

We may (in principle) build robots of any kind; but, whatever their architecture, behavior, etc., may be, they can never be conscious. This is not a principled position, of course: consciousness is a matter of biology, not of computations; and that it cannot be a property of computational systems has been argued by, among others, Bringsjord (1992, 1994) and Searle (1992). Thus, there is no reason to expect artificial consciousness, unless we can duplicate the relevant properties of biological nervous tissue (connectionism, of course, offers no solution in this respect, since artificial neural nets are exactly as neural as so-called classical computational systems are).

Thus, we have at least one property of biological systems that cannot be duplicated in robots. Does this make any difference as to how the respective control systems work? Three positions

may be taken with respect to the role of consciousness in the control systems of high-level adaptive agents: consciousness either [a] has nothing whatsoever to do with control, or it has; in the latter case, it may be either [b] a necessary feature or [c] a contingent one. In the following, I will briefly discuss these three positions. While position [a] seems unsustainable, there currently seems to be no reason to choose between positions [b] and [c].

## Position [a]: Consciousness is irrelevant to control

According to position [a], consciousness exerts no role at all in the control of our behavior, which is assured by nonconscious machinery. In spite of its widespread diffusion in cognitive science, this position is quite unreasonable, at least because of evolutionary concerns.

Human beings have a first-person (conscious) understanding of their own behavior as being, at least in part, guided by conscious deliberation. They also tend to interpret other individuals' behavior in the same perspective. This might be their mistake, of course: behavior might be completely independent of consciousness, which would then come only as a post hoc explanation to oneself of what is actually a sterile byproduct of one's nonconscious machinery. Hard as it may be to accept this idea, it is harder to prove it wrong.

Such a radical position is actually a common one in cognitive science, where it is usual to conceive of the mind as a bunch of boxes "processing" a flow of input information in order to determine an output. Since this is thought to be sufficient to produce intelligent behavior, consciousness is *de facto* perceived as irrelevant to the human kind.

It follows that either consciousness does not exist (which, as I can personally guarantee, is not the case), or it is superfluous. In the latter case, biological minds must be something more than control systems, because they exhibit a property which is completely unrelated to control. But then, under what selective pressure might such a property have evolved? It should be noticed that social behavior is a matter of control indeed. This means, first, that something completely unrelated to control can play no role in mating; and, second, that problems of control do not concern only low-level, possibly nonconscious, processes.

Thus, it would be quite amazing if consciousness turned out to be an extravagant byproduct of our brains. If this were the case, however, appropriately designed robots would be able to pass any version of the Turing Test, no matter how refined (Turing, 1950; Harnad, 1991): consciousness is a first-person property; were it a useless one too, how could we ever know whether the individual sitting in front of us was a robot?

## Position [b]: Consciousness is necessary for (high-level) control

At the other end of the continuum is the idea that consciousness is a necessary feature of high-level control systems (position [b]). In this perspective, our being conscious, far from being an extravagant luxury, would be intimately connected to the distinctive behavioral performance exhibited by our species (and possibly by others as well).

This would be in accordance with evolutionary concerns: complex-minded species need to be conscious, and that's all. The question of whether this pertains only to the human species, as well as whether there are middle steps between the nonconscious and the conscious, should be viewed as empirical matters, to be settled with yet-to-be-devised methodologies. In this case, no robot, however accurately designed, could ever pass a refined version of the Turing Test: some predictable detail, though possibly an imperceptible one, would be guaranteed to betray the artificial.

The problem here is that we do not have the slightest idea of what consciousness is for. As soon as we have a theory of the role it might play in some mental process, we are *ipso facto* excluding it from that very process, thereby restricting it to all the processes we have not yet understood. This does not imply that consciousness does not exist. Rather, consciousness is a first-person property, whereas scientific theories must be stated in the third person: if we could devise a third-person theory of consciousness, we would simply rule out the need for con-

sciousness itself. The problem is, of course, that as a matter of fact human beings are conscious.

This is not meant to imply that we cannot theorize about consciousness: theories of consciousness might be possible in principle, but they may require some major shift in our conception of what a cognitive theory is.

## Position [c]: Consciousness is relevant but not necessary to control

According to position [c], consciousness is relevant to control, but not necessary. We must acknowledge, as a matter of fact, its crucial role in the control of our behavior, but we cannot exclude the possibility that analogous results may be achieved with different methods.

In this perspective, the role of consciousness is restricted to biological agents, leaving room for computational ones which be unconscious but able to behave in a human-like fashion. There might exist nonconscious forms of control which work equally well, and with exactly the same results, in producing high-level behavior. The difference would be that humans entertain conscious, first-person mental states, whereas robots should happily limit themselves to third-person computations over mental states equivalents (whatever that might mean). This is Bringsjord's <ROB> thesis (Bringsjord, 1992), as I understand it.

Such a position rules out neither the need for suitable theories of biological consciousness (i.e., the need for a change in the postulates underlying the psychological branches of cognitive science), nor the possibility for robots to pass however refined versions of the Turing Test.

In spite of my personal sympathy for this position, I can adduce no empirical support in its favor. An empirical, though obviously weak, case *against* it might be found in the relatively poor accomplishments of current AI, if viewed from the psychologist's viewpoint: no existing system can be said to reach the level of complexity of even an insect. The hope is that this is not the symptom of the difficulties I have described under position [b]. As is common in science, only time will tell.

## Conclusions

Both positions [b] and [c] imply a separation between AI and psychology; the cleavage, however, would follow different lines. In the [b] case, robots could never aspire to great accomplishments: they would be limited to low-level behaviors, such as those exhibited by presumably nonconscious animals (insects? reptiles?). Accordingly, relationships should be found between AI and ethology rather than psychology or animal cognition.

In the [c] case, there would be no principled limits to robots' possibilities; there might be relationships between AI and psychology, though far from those usually conceived. Rather than considering the similarities in the computations carried out, as is usually recommended (see, e.g., Pylyshyn, 1984), it might be more interesting to study issues like architectural requirements for efficient control, requirements on initial knowledge, etc. Once we have accepted this idea, the quest for Turing-testable robots becomes meaningless: Turing-testable behaviors are those of our species, and there is no (scientific) point in trying to simulate them when we know that computational consciousness is impossible. Different species will have different abilities, so why not disconnect Turing machines from the Turing test and study robots as an evolutionary line independent from ours?

## References

Bringsjord, S. (1992) *What robots can and can't be*. Boston: Kluwer Academic.
Bringsjord, S. (1994) Précis of: What robots can and can't be. *PSYCOLOQUY* 5(59) robot-consciousness.1. bringsjord.

Harnad, S. (1991) Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines* 1:43-54.

Newell, A. (1990) *Unified theories of cognition.* Boston: Harvard University Press.

Pylyshyn, Z.W. (1984) *Computation and cognition.* Boston: MIT Press.

Searle, J.R. (1992) *The rediscovery of the mind.* Boston: MIT Press.

Turing, A.M. (1950) Computing machinery and intelligence. *Mind* 59:433-460.