

RECONCILING DESCRIPTIONS OF CONSCIOUSNESS
FROM WITHIN AND FROM WITHOUT

In its attempt to come to terms with consciousness or the knowing subject, Western philosophy has been characterised by a tension between two points of view. Consciousness has been described from “within”, as in Husserl’s *Transcendental Phenomenology*, or it has been described from “without,” as it is in Behaviourism and Artificial Intelligence. While Husserl has tried to reduce all “externality” to whatever it is for the transcendental consciousness, others have tried to reduce all “internality” to something to be found in the external world, like behaviour or neurological processes. Merleau-Ponty has consistently argued against this dualism, claiming that we do not have to choose between a philosophy which “takes our experience from ‘within’ and a philosophy that would judge it from without.” (1968:160) Inside and outside, he says, are inseparable.¹

The objective of this essay is to show that in their reductionist arguments the advocates of Artificial Intelligence (A.I.) exploit, without recognising it or acknowledging it, another point of view, one which transcends the dualism of the internal and the external points of view. We will show that if this “other” point of view were acknowledged, however, the fundamental arguments on which this approach to consciousness is based would collapse.

CONSCIOUSNESS FROM WITHIN

Reflecting on consciousness from within, I find it impossible to assume things about myself which others, approaching me from the “outside”, are able to assume about me. The demands of internal consistency makes it impossible for me to recognise with the advocates of A.I., the possibility that all my thoughts are the effects of the physical and chemical properties of my nervous system. Internal consistency dictates that in adopting such a theory of the mind I would be led into a vicious circle, where I would have to concede that I could never know what it was that I had accepted, or have any good reason for accepting it.

If I were to accept the possibility that all my thoughts are the effects of causes, if I were to accept, for example, that I was led to conclusions, not because they followed on their premises, but that these conclusions came to

mind because of the physical and chemical properties of my brain, I would also have to accept the possibility that I was not rational and hence that I would never be able to distinguish between valid and invalid arguments. Let us take as an example the following syllogism: If "A implies B" and "B implies C" then "A implies C." If all my thoughts were the effects of causes, I would "think" or "reach" the conclusion "A implies C," because I had been programmed to think "A implies C" whenever I was presented with "A implies B" and "B implies C." But of course I could have been incorrectly programmed. In which case it could be that the premises do not actually imply the conclusion, and I would continue drawing false conclusions, without ever being able to discover that they were false.

It could be argued that, like any computer, I could have a verifying function which would verify the conclusion which I had been caused to think. But since the "result" of the verifying function would itself be the effect of causes, it could itself be defective, I would still never know whether any of my conclusions are valid. Furthermore, my knowledge that I have such a verifying function could only reassure me that I was being led to the correct conclusion, if I already knew that having a verifying function *implies* that my conclusions are most probably correct, but I could only *know* this if I already knew that what I took to be implied by having a verifying function, actually was implied by having such a verifying function. In other words I already need to know that I have been correctly programmed.

Similarly, it has been argued that I could accept that all my thoughts are the effects of causes and still remain confident that I have been correctly programmed, because if I were not, I would not have survived. But arguments like these could never provide me with good reasons for believing that I was rational or correctly programmed. These arguments, like any argument, should only be accepted if they are valid, and I could only trust my identification of their validity if I knew that I was rational. I need to be sure, therefore, that I am rational or correctly programmed before I can accept the argument. Similarly, I can never convince myself that I am rational by applying some test or other of rationality, because I need to be rational to be able to evaluate the test and to give a reliable interpretation of its results.

Every act of verification *presupposes* that my powers of judgment lead me to valid conclusions. How could I establish anything *independently* of my own thinking? Nor could I use the judgements of others as a guide to verify my thinking, for whether or not my thinking *conforms to* or *contradicts* their thinking is something I would still have to establish through my own judgements and hence by assuming that I was rational. For me there could be

no proof or evidence that I was rational, for I need to know that I am rational before accepting any argument or evidence for my rationality. This means that I could only accept that all my thoughts are the effects of the physical and chemical properties of my nervous system, *in an act of blind faith*. I could never *know* that such an assumption was reasonable, nor could I ever *know* what such an assumption entailed.

Whatever conception of the mind I may wish to *justify*, the conception must be compatible with my assumption that I am able to *justify* a conception of the mind. If I reflect on myself at the very moment that I carry out such a justification, at the moment, for example, that I “draw” a conclusion from premises, I find that the conclusion is not for me the effect of causes, it is not for me as if I have been programmed such that under certain conditions a conclusion simply “comes to mind,” appears to me without me knowing why this one and not another. On the contrary, I actively and self-consciously “pursue” *the* conclusion itself as that which is implied by the premises, rather than being passive to whatever “conclusion” the chemistry of my brain happens to lead me. From my point of view, my mind is such that I *think* “A implies C” because given the premises, “A” *actually does* imply “C.” In my thinking I direct myself towards that conclusion which is actually implied by the premises, rather than allowing my thinking to be determined by the properties of my brain. I do not have to prove to myself that my thinking is rational, because in pursuing “the” conclusion I *make myself rational*.

For me to “pursue” the conclusion, to “pursue” that which is implied by the premises, is not to be guided by rules, images or intuitions. I could never equate my rationality with following rules, such as the rules of logic, for I need to be rational in order to follow rules correctly. If being rational was always reducible to following rules, I would then need rules for following rules, and so on. For the same reason I could not be guided by intuitions and images of the valid conclusion, for how would I ever know that the appropriate images or intuitions have been conjured up, and that I have made the appropriate use of them. For the same reason, my thinking is not guided by my knowledge. My thinking does not move to “Socrates is rational” because I *know* that this is implied by the premises, “All men are rational” and “Socrates is a man.” *I think in order to know*. My awareness of the conclusion cannot play any role in my thinking, because it has not yet been inferred, and although in actual fact it *is* implied by the premises, it is not yet the conclusion *for me*.

From my point of view, I avoid the vicious circle because for me my thought is intentional, i.e., it is able to pursue a truth which is not relative to

my neurological state, or relative to any set of rules or formulae for thinking. I am able to pursue that for which I possess no formula or representation in advance, without there being in me another mind which already has the answer and which could direct my thinking, or which could certify that I had arrived at the correct conclusion. I avoid the viscous circle because I can direct myself to that which is not present to me in any form. At the moment that I understand the premises, the conclusion, as that which follows, becomes "something to be grasped in thought." I am "aware" of it only as I am aware of objects which are behind my back, as something "to be reached," as the conclusion "to be disclosed." The disclosure occurs when a certain voluntary attitude I take up suddenly receives from outside the confirmation for which it was waiting. I aim at something which for me is no more than that which is implied by the premises, and suddenly the conclusion takes possession of my thought, and I surrender myself to it, such that whatever thought processes are necessary for it to become an object of thought will be elicited by it as it discloses itself to me, or as it "comes into view"² (Merleau-Ponty, 1962:212).

In the experience of trying to recall the name of an actor or a president, which is on the "tip of the tongue," we can get some idea of the ability of the mind to be directed to something without rules for finding the name or without images of the name. Yet in trying to recall a name, my thinking is not haphazard or random, and it is not through a process of trial and error that I finally arrive at the name, and that I need to *infer* that this was the name for which I was searching. For me my thought is nonrandom because it is intentional, because it is directed beyond any contents of consciousness, beyond any representation and the only explanation for the direction of my thought is the name itself which it attempts to grasp.

This experience of the intentionality of our own thinking is reflected in the reports of most thinkers who have tried to describe their own thinking. In his *Meditations*, Descartes, for example, accepts that his thinking is intentional without subjecting his experience of his thinking to methodic doubt. When he arrives at his "Cogito Ergo Sum," he never considers the possibility that all that he has done is reveal some peculiarity of his mind, some subjective tendency of his thinking, such that whenever he thinks about the fact that he is thinking, the thought that he exists always appears. He never considers the possibility that even though his thought "Sum" always follows his thought "Cogito," the first thought may simply cause the second, and that the fact that he thinks might not actually imply that he exists. Nor could he claim that "Sum" always follows "Cogito" *according to the rules of thought*, for then he

would have to concede that “Cogito Ergo Sum” would be true only if the rules of thought were reliable, which would be the condition of his conclusion if he thought of himself as a computer. If for Descartes his own thinking could be nonrandom only if it was guided by rules, he would not have claimed that the “Cogito Ergo Sum” was indubitable and could serve as his Archimedean point. The truth of the Cogito would itself depend on the reliability of the rules. Descartes avoids an infinite regress because it is for him beyond question that his thinking is, as he experiences it, nonrandom, simply because it is intentional. He has no hesitation in accepting that he is able to conclude “Sum” without following rules or being guided by images, and hence that there can be no other explanation for his concluding “Sum” than the fact that his existence is *actually* implied by the fact that he thinks.

This is how I experience the power of my own thought to transcend its neurological base, and this is why for me, at the moment that I draw a conclusion, it is essential that my thinking is in itself as it presents itself to me. If I were to concede the possibility that my real thinking processes were numerically distinct from what I experience myself as doing, I could have no confidence in the validity of my reasoning, and hence no way of establishing the “objectivity” of my experience.

THE VIEW ON CONSCIOUSNESS FROM THE OUTSIDE

However, no matter what I may or may not assume about myself, isn't it always possible for someone else to assume that all my experiences, even the experience of “reaching out” towards a truth which is not relative to my neurological makeup, is no more than a subjective impression? From such an external point of view there is no evidence that I have any “relationship” with the conclusion “itself,” that my thinking could ever transcend its neurological base. From such an external view there are only neurological events, explicable in terms of chemical and physical causes, and it is inconceivable that “the” conclusion of the premises, or “the” truth itself, could play any direct role in the processes. Consequently isn't she always able to assume that all my conscious processes are dependent on neurological ones, and that I will always be limited to what my nervous system can produce? Or on the other hand, couldn't she assume that the processes through which I follow my “laws of logic” or am guided by images and intuitions, are unconscious, and that this experience I have of “prevailing over my own subjectivity,” of reaching a truth which is not relative to *my* logic and *my* use of it, is an illusion? Couldn't she argue that although *I* can't make these assumptions