of the members of the subset to the truth of the members of the remainder.[9]

One source of resistance to the idea that science can and ought to help to determine philosophical truth may be this. It is often assumed, and may be true, that the discoveries of science are contingent truths, while conceptual analysis issues in necessary truths. Now there is an old tradition in philosophy that links necessary truth with self-evidence, or logical indubitability, while contingent truth is linked with lack of self-evidence, or the possibility of error. It is easy to show, and now widely appreciated, that there is in fact no such link. Mistake about any moderately complex logical necessity is quite easy. Contrariwise, the class of truths for which contemporary philosophy most frequently claims indubitability are first-person reports about current mental states, propositions that are certainly not necessary truths. Nevertheless, the old linking persists in philosophers' thoughts, even though it is recognized to be an error. From the perspective of the search for certainty, necessity is thought to be an *epistemologically* desirable characteristic. And so there is resistance to the idea that the contingent truths of science could be any guide to the conceptual truths of philosophical analysis.

But whatever may be the epistemological credit of the propositions put forward by logicians and mathematicians, the conceptual analyses of philosophers are matters of the greatest dispute among those most competent to judge (the philosophers themselves), as well as among everybody else. The rational man therefore will not give them a high epistemological credit rating. He will recognize that philosophy needs all the help it can get in deciding upon the truth or falsity of conceptual claims. He therefore will take the most careful notice of any relevant relative certainties that science or common sense can provide.

9. A good illustration of what I am arguing for is provided by Keith Gundersen (1971: 92), when he says: "I want to suggest that by trying to program a machine with recognition capacities we will perhaps improve our understanding of the concept of recognition."

4
What is Consciousness?

Proc. of the Russellian Soc.
(Sydney, 1978)
Repr. in D.M. Armstrong, The Nature of Mind & Other Essays (1980)

The notion of consciousness is notoriously obscure. It is difficult to analyze, and some philosophers and others have thought it unanalysable. It is not even clear that the word "consciousness" stands for just one sort of entity, quality, process, or whatever. There is, however, one thesis about consciousness that I believe can be confidently rejected: Descartes' doctrine that consciousness is the essence of mentality. That view assumes that we can explain mentality in terms of consciousness. I think that the truth is in fact the other way round. Indeed, in the most interesting sense of the word "consciousness", consciousness is the cream on the cake of mentality, a special and sophisticated development of mentality. It is not the cake itself. In what follows, I develop an anti-Cartesian account of consciousness.

MINIMAL CONSCIOUSNESS

In thinking about consciousness, it is helpful to begin at the other end and consider a totally unconscious person. Somebody in a sound, dreamless sleep may be taken as an example. It has been disputed whether unconsciousness is really ever total. There is some empirical evidence that a person in dreamless sleep, or even under a total anaesthetic, still has some minimal awareness. Minimal behavioural reactions to sensory stimuli have been observed under these conditions. But let us take it, if only as a simplifying and perhaps unrealistic assumption, that we are dealing with *total* unconsciousness.

Notice first that we are perfectly happy to concede that such a person, while in this state of total unconsciousness, has a *mind*. Furthermore, although by hypothesis this mind is in no way active — no mental events take place, no mental processes occur within it — we freely allow that this mind is in various *states*.

The totally unconscious person does not lack knowledge and beliefs. Suppose him to be a historian of the mediaeval period. We will not deny him a great deal of knowledge of and beliefs about the Middle Ages just because he is sound asleep. He cannot give current expression to his knowledge and his beliefs, but he does not lack them. The totally unconscious person also may be credited with memories. He also can be said to have skills, including purely mental skills such as an ability for mental arithmetic. The ability is not lost during sound sleep just because it then cannot be exercised, any more than an athlete loses his athletic abilities during sound sleep, when he cannot exercise them. A totally unconscious person may be credited with likes and dislikes, attitudes and emotions, current desires and current aims and purposes. He may be said to have certain traits of character and termperament. He may be said to be in certain moods: "He has been depressed all this week."

How are we to conceive of these mental states (it seems natural to call them "states") we attribute to the unconscious person? Some decades ago, under the influence of positivistic and phenomentalistic modes of thought, such attributions of mental states to an unconscious person would not have been taken very seriously, ontologically. It would have been thought that to say that the currently unconscious person *A* believes that *p*, is simply to refer to various ways in which *A* 's mind works, or would work in suitable circumstances, before and/or after he wakes up. (The same positivist spirit might try further to reduce the way that *A* 's mind works to *A* 's peripheral bodily behaviour or to the behaviour *A* would exhibit in suitable circumstances.)

In historical perspective, we can see clearly how unsatisfactory such a view is. Consider two persons, *A* and *B*, unconscious at the same time, where it is true of *A* that he believes that *p*, but false of *B*. Must there not be a difference between *A* and *B* at that time to constitute this difference in belief-state? What else in the world could act as a truth-maker (the ground in the world) for the different conditional statements that are true of *A* and *B* ? The mind of the unconscious person cannot be dissolved into statements about what would be true of the person *if* the situation were other than it was; if, in particular, he were not unconscious.

In considering this point, I find very helpful the analogy between an unconscious person and a computer that has been programmed in various ways, that perhaps has partially worked through certain routines and is ready to continue with them, but is not currently operating. (I do not think that anything in the analogy turns on the material, physical nature of the computer. Even if the mind has to be conceived of in some immaterial way, the analogy will still hold.) The computer, perhaps, will have a certain amount of information stored in its memory-banks. This stored information may be compared to the knowledge, belief and memories the unconscious person still has during unconsciousness. If a Materialist account of the mind is correct, then, of course, knowledge, belief and memory will be physically encoded in the brain in some broadly similar way to the way in which information is stored in the computer. But the Dualist, say, will equally require the conception of immaterial storage of knowledge, belief and memory.

What we can say both of the knowledge, beliefs, etc. possessed by the totally unconscious person, and also of the information stored in the switched-off computer, is that they are *causally quiescent*. Of course, nothing is causally quiescent absolutely: while a thing exists, it has effects upon its environment. But the information stored in the switched-off computer is causally quiescent with respect to the computing operations of the computer, and for our purposes this may be called causal quiescence. (The information may remain causally quiescent even after the computer has been switched on, unless that piece of information is required for current calculations.) In the same way, knowledge and beliefs may be said to be causally quiescent while they are not producing any *mental* effect in the person. The mental states of a totally unconscious person are thus causally quiescent (if they are not, we may stipulate that the person is not totally unconscious). Knowledge, beliefs, and so on may remain causally quiescent in

this sense even when the mind is operational, for instance, where there is no call to use a particular piece of knowledge.

It seems, then, that we attribute mental states of various sorts to a totally unconscious person. But there are certain mental attributions we do not make. The totally unconscious person does not perceive, has no sensations, feelings or pangs of desire. He cannot think, contemplate or engage in any sort of deliberation. (He can have purposes, because purposes are capable of causal quiescence, but he cannot be engaged in carrying them out.) This is because perception, sensation and thinking are mental *activities* in a way that knowledge and beliefs are not. The distinction appears, roughly at any rate, to be the distinction between events and occurrences on the one hand, and states on the other. When a mental state is producing mental effects, the comings-to-be of such effects are mental events: and so mental activity is involved.

We now have a first sense for the word "consciousness". If there is mental activity occurring in the mind, if something mental is actually happening, then that mind is not totally unconscious. It is therefore conscious. A single faint sensation is not much, but if it occurs, to that extent there is consciousness. Unconsciousness is not total. I call consciousness in this sense "minimal" consciousness.

It is alleged that it sometimes occurs that someone wakes up knowing the solution to, say, a mathematical problem, which they did not know when they went to sleep. If we rule out magical explanations, then there must have been mental activity during sleep. To that extent, there was minimal consciousness. This is compatible with the completest "unconsciousness" in a sense still to be identified.

PERCEPTUAL CONSCIOUSNESS

Among the mental activities, however, it appears that we make a special link between consciousness and *perception*. In perception, there is consciousness of what is currently going on in one's environment and in one's body. (Of course, the consciousness may involve illusion.) There is an important sense in which, if a person is not perceiving, then he is not conscious, but if he is perceiving, then he is conscious. Suppose somebody to be dreaming. Since there is mental activity going on, the person is not totally unconscious. He is minimally conscious. Yet is there not some obvious sense in which he is unconscious? Now suppose that this person starts to perceive his environment and bodily state. (I do not want to say "suppose he wakes up", because perhaps there is more to waking up than just starting to perceive again.) I think that we would be inclined to say that the person was now conscious in a way that he had not been before, while merely dreaming. Let us say, therefore, that he has regained "perceptual" consciousness. This is a second sense of the word "consciousness". Perceptual consciousness entails minimal consciousness, but minimal consciousness does not entail perceptual consciousness.

INTROSPECTIVE CONSCIOUSNESS

Let us suppose, now, that there is mental activity going on in a person, and that this activity includes perception. If what has been said so far is accepted, then there are two senses in which such a person can be said to be conscious. He or she has *minimal* consciousness and has *perceptual* consciousness. There is, nevertheless, a third sense, in which such a person may *still* "lack consciousness". Various cases may be mentioned here. My own favourite is the case of the long-distance truck-driver. It has the advantage that many people have experienced the phenomenon.

After driving for long periods of time, particularly at night, it is possible to "come to" and realize that for some time past one has been driving without being aware of what one has been doing. The coming-to is an alarming experience. It is natural to describe what went on before one came to by saying that during that time one lacked consciousness. Yet it seems clear that, in the two senses of the word that we have so far isolated, consciousness was present. There was mental activity, and as part of that mental activity, there was perception. That is to say, there was minimal consciousness and perceptual consciousness. If there is an inclination to doubt this, then consider the extraordinary sophistication of the activities successfully undertaken during the period of "unconsciousness".

A purpose was successfully advanced during that time: that of driving a car along a road. This purpose demanded that various complex sub-routines be carried out, and carried out at appropriate points (for instance, perhaps the brake or the clutch was used). Were not these acts purposeful? Above all, how is it possible to drive a car for kilometres along a road if one cannot perceive that road? One must be able to see where one is going, in order to adjust appropriately. It would have to be admitted, at the very least, that in such a case, eyes and brain have to be stimulated in just the same way as they are in ordinary cases of perception. Why then deny that perception takes place? So it seems that minimal consciousness and perceptual consciousness are present. But something else is lacking: consciousness in the most interesting sense of the word.

The case of the long-distance truck-driver appears to be a very special and spectacular one. In fact, however, I think it presents us with what is a relatively simple, and in evolutionary terms relatively primitive, level of mental functioning. Here we have more or less skilled purposive action, guided by perception, but apparently no other mental activity, and in particular no consciousness in some sense of "consciousness", which differs from minimal and perceptual consciousness. It is natural to surmise that such relatively simple sorts of mental functioning came early in the course of evolutionary development. I imagine that many animals, particularly those whose central nervous system is less developed than ours, are continually, or at least normally, in the state in which the long-distance truck-driver is in temporarily. The third sort of consciousness, I surmise, is a late evolutionary development.

What is it that the long-distance truck-driver lacks? I think it is an additional form of perception, or, a little more cautiously, it is something that resembles perception. But unlike *sense*-perception, it is not directed towards our current environment and/or our current bodily state. It is perception of the mental. Such "inner" perception is traditionally called introspection, or introspective awareness. We may therefore call this third sort of consciousness "introspective" consciousness. It entails minimal consciousness. If perceptual consciousness is restricted to sense-perception, then

introspective consciousness does not entail perceptual consciousness.

Introspective consciousness, then, is a perception-like awareness of current states and activities in our own mind. The current activities will include sense-perception: which latter is the awareness of current states and activities of our environment and our body. And (an important and interesting complication) since introspection is itself a mental activity, it too may become the object of introspective awareness.

Sense-perception is not a *total* awareness of the current states and activities of our environment and body. In the same way, introspective consciousness is not a total awareness of the current states and activities of our mind. At any time there will be states and activities of our mind of which we are not introspectively aware. These states and activities may be said to be unconscious mental states and activities in one good sense of the word "unconscious". (It is close to the Freudian sense, but there is no need to maintain that it always involves the mechanism of repression.) Such unconscious mental states and activities of course may involve minimal and/or perceptual consciousness, indeed the *activities* involve minimal consciousness by definition.

Just as perception is selective — not all-embracing — so it also may be mistaken. Perceptions may fail to correspond, more or less radically, to reality. In the same way, introspective consciousness may fail to correspond, more or less radically, to the mental reality of which it is a consciousness. (The indubitability of consciousness is a Cartesian myth, which has been an enemy of progress in philosophy and psychology.)

Following Locke, Kant spoke of introspection as "inner sense", and it is essentially Kant's view I am defending here. By "outer sense", Kant understood sense-perception. There is, however, one particular form of "outer sense" that bears a particularly close formal resemblance to introspection. This is bodily perception or *proprioception*, the perception of our own current bodily states and activities. If we consider the objects of sight, sound, touch, taste and smell, then we notice that such objects are intersubjectively available. Each of us is capable of seeing or touching numerically the very same physical surface, hearing numerically

the very same sound, tasting numerically the same tastes or smelling numerically the same smell. But the objects of proprioception are not intersubjectively available in this way.

Consider, for instance, kinaesthetic perception, which is one mode of proprioception. Each person kinaesthetically perceives (or, in some unusual cases, misperceives) the motion of his own limbs and those of nobody else. There is no overlap of kinaesthetic objects. This serves as a good model for, and at the same time it seems to demystify, the privacy of the objects of introspection. Each of us perceives current states and activities in our own mind and that of nobody else. The privacy is simply a little more complete than in the kinaesthetic case. There are other ways to perceive the motion of my limbs besides kinaesthetic perception — for instance, by seeing and touching. These other ways are intersubjective. But, by contrast, nobody else can have the direct awareness of my mental states and activities that I have. This privacy, however, is contingent only. We can imagine that somebody else should have the same direct consciousness of my mental states and activities that I enjoy. (They would not *have* those states, but they would be directly aware of them.)

Perception is a causal affair. If somebody perceives something, then it is involved in the perception; it is even involved in the concept of perception: that the thing perceived acts upon the perceiver, causing the perception of the object. If introspective consciousness is to be compared to perception, then it will be natural to say that the mental objects of introspection act within our mind so as to produce our introspective awareness of these states. Indeed, it is not easy to see what other naturalistic account of the coming-to-be of introspections could be given. If introspection is a causal process, then it will follow, incidentally, from our earlier definition of causal quiescence that whenever we are introspectively aware of one of our mental states, then that state is not at that time causally quiescent.

TYPES OF INTROSPECTIVE CONSCIOUSNESS

Perhaps we still have not drawn enough distinctions. Sometimes the distinction is drawn between mere "reflex" consciousness,

which is normally always present while we are awake (but which is lost by the long-distance truck-driver), and consciousness of a more explicit, self-conscious sort.

This difference appears to be parallel to the difference between mere "reflex" seeing, which is always going on while we are awake and our eyes are open, and the careful *scrutinizing* of the visual environment that may be undertaken in the interest of some purpose we have. The eyes have a watching brief at all times that we are awake and have our eyes open; in special circumstances, they are used in a more attentive manner. (In close scrutiny by human beings, introspective consciousness is often, although not invariably, also called into play. We not only give the object more attention but have a heightened awareness of so doing. But, presumably, in lower animals such attentive scrutiny does not have this accompaniment.) Similarly, introspective consciousness normally has only a watching brief with respect to our mental states. Only sometimes do we carefully scrutinize our own current state of mind. We can mark the distinction by speaking of "reflex" introspective awareness and opposing it to "introspection proper". It is a plausible hypothesis that the latter will normally involve not only introspective awareness of mental states and activities but also introspective awareness of that introspective awareness. It is in any case a peculiarly sophisticated sort of mental process.

WHAT IS SO SPECIAL ABOUT INTROSPECTIVE CONSCIOUSNESS?

There remains the feeling that there is something quite special about introspective consciousness. The long-distance truck-driver has minimal and perceptual consciousness. But there is an important sense, we are inclined to think, in which he has no experiences, indeed is not really a person, during his period of introspective unconsciousness. Introspective consciousness seems like a light switched on, which illuminates utter darkness. It has seemed to many that with consciousness in this sense, a wholly new thing enters the universe.

I now will attempt to explain why introspective consciousness *seems to have*, but does not necessarily *actually* have, a quite special status in the world. I proceed by calling attention to two

points, which will then be brought together at the end of the section.

First, it appears that introspective consciousness is bound up in a quite special way with consciousness of self. I do not mean that the self is one of the particular objects of introspective awareness alongside our mental states and activities. This view was somewhat tentatively put forward by Russell in *The Problems of Philosophy* (1912: Ch. 5), but had already been rejected by Hume and by Kant. It involves accepting the extraordinary view that what seems most inward to us, our mental states and activities, are not really us. What I mean rather is that we take the states and activities of which we are introspectively aware to be states and activities of a single continuing thing.

In recent years, we have often been reminded, indeed admonished, that there is a great deal of theory involved even in quite unsophisticated perceptual judgements. To see that there is a tomato before our body is already to go well beyond anything that can be said to be "given", even where we do not make excessive demands (such as indubitability) upon the notion of the given. Consider knowingly perceiving a tomato. A tomato, to be a tomato, must have sides and back, top and bottom, a certain history, certain casual powers; and these things certainly do not seem to be given in perception. If we consider the causal situation, it is only the shape, size and colour of some portions of the surface of the tomato (the facing portions) that actually determine the nature of the stimulation that reaches our eyes. This suggests that, at best, it is only these properties that are in any way "given" to us. The rest is, in some sense, a matter of theory, although I do not think that we should take this to mean that the perceptual judgement that there is a tomato before us is a piece of risky speculation.

It is therefore natural to assume that the perceptions of "inner sense" involve theory, involve going beyond the "given", in the same general way that the perceptions of "outer sense" do. In particular, whatever may be the case with other animals, or with small children, or with those who, like the Wild Boy of Aveyron, have not been socialized, for ordinary persons, their mental states and

activities are introspected as the states and activities of a single thing.

Once again, the comparison with proprioception seems to be instructive. We learn to organize our proprioceptions so that they yield us perceptions of a single, unitary, physical object, our body, concerning which our proprioceptions give us certain information: its current posture, temperature, the movement of its limbs, and so on. This is clearly a theoretical achievement of some sophistication.

In the same way, we learn to organize what we introspect as being states of, and activities in, a single continuing entity: our self. Mere introspective consciousness, of course, is not at all clear just what this self is. At a primitive level perhaps, no distinction is made between the self and the body. Identification of the thing that is introspected as, say, a spiritual substance, or as the central nervous system, goes far beyond the level of theorizing involved in ordinary introspection. But the idea that the states and activities observed are states and activities of a unitary thing *is* involved. Introspective consciousness is consciousness of self.

If it is asked why introspection is theory-laden in this particular way, then an answer can be suggested. It is always worth asking the question about any human or animal organ or capacity: "What is its biological function?" It is therefore worth asking what is the biological function of introspective consciousness. Once the question is asked, then the answer is fairly obvious: it is to sophisticate our mental process in the interests of more sophisticated action.

Inner perception makes the sophistication of our mental processes possible in the following way. If we have a faculty that can make us aware of current mental states and activities, then it will be much easier to achieve *integration* of the states and activities, to get them working together in the complex and sophisticated ways necessary to achieve complex and sophisticated ends.

Current computer technology provides an analogy, though I would stress that it is no more than an analogy. In any complex computing operation, many different processes must go forward simultaneously: in parallel. There is need, therefore, for an overall plan for these activities, so that they are properly co-ordinated. This cannot be done simply in the manner in which a "command

economy" is supposed to be run: by a series of instructions from above. The co-ordination can only be achieved if the portion of the computing space made available for administering the overall plan is continuously made "aware" of the current mental state of play with respect to the lower-level operations that are running in parallel. Only with this feedback is control possible. Equally, introspective consciousness provides the feedback (of a far more sophisticated sort than anything available in current computer technology) in the mind that enables "parallel processes" in our mind to be integrated in a way that they could not be integrated otherwise. It is no accident that fully alert introspective consciousness characteristically arises in *problem* situations, situations that standard routines cannot carry one through.

We now can understand why introspection so naturally gives rise to the notion of the self. If introspective consciousness is the instrument of mental integration, then it is natural that what is perceived by that consciousness should be assumed to be something unitary.

There is nothing necessary about the assumption. It may even be denied on occasion. Less sophisticated persons than ourselves, on becoming aware of a murderous impulse springing up, may attribute it not to a hitherto unacknowledged and even dissociated part of themselves, but to a devil who has entered them. In Dickens' *Hard Times*, the dying Mrs Gradgrind says that there seems to be a pain in the room, but she is not prepared to say that it is actually *she* that has got it. In her weakened condition, she has lost her grip upon the idea that whatever she introspects is a state of one unitary thing: herself.

But although the assumption of unity is not necessary, it is one we have good reason to think true. A Physicalist, in particular, will take the states and activities introspected to be all physical states and activities of a continuing physical object: a brain.

That concludes the first step in my argument: to show that, and in what sense, introspective awareness is introspective awareness of self. The second step is to call attention to the special connection between introspective consciousness and event-memory, that is, memory of individual happenings. When the long-distance truck-driver recovers introspective consciousness, he has no

memory of what happened while it was lacking. One sort of memory-processing cannot have failed him. His successful navigation of his vehicle depended upon him being able to *recognize* various things for what they were and treat them accordingly. He must have been able to recognize a certain degree of curve in the road, a certain degree of pressure on the accelerator, for what they were. But the things that happened to him during introspective unconsciousness were not stored in his event-memory. He lived solely in the present.

It is tempting to suppose, therefore, as a psychological hypothesis, that unless mental activity is monitored by introspective consciousness, then it is not remembered to have occurred, or at least it is unlikely that it will be remembered. It is obvious that introspective consciousness is not sufficient for event-memory. But perhaps it is necessary, or at least generally necessary. It is notoriously difficult, for instance, to remember dreams, and it is clear that, in almost all dreaming, introspective consciousness is either absent or is at a low ebb.

So it may be that introspective consciousness is essential or nearly essential for event-memory, that is, memory of the past as past. *A fortiori*, it will be essential or nearly essential for memory of the past of the self.

The two parts of the argument now may be brought together. If introspective consciousness involves (in reasonably mature human beings) consciousness of self, and if without introspective consciousness there would be little or no memory of the past history of the self, the apparent special illumination and power of introspective consciousness is explained. Without introspective consciousness, we would not be aware that we existed — our self would not be self to itself. Nor would we be aware of what the particular history of that self had been, even its very recent history. Now add just one more premiss: the overwhelming interest that human beings have in themselves. We can then understand why introspective consciousness can come to seem a condition of anything mental existing, or even of anything existing at all.