
CONSCIOUSNESS: THE TRANSMUTATION OF A CONCEPT

BY

PATRICIA SMITH CHURCHLAND

INTRODUCTION*

It sometimes happens in the history of science that well-used, highly entrenched, revered and respected concepts come unstuck. That is, under the suasion of a variety of empirical-cum-theoretical forces, certain concepts lose their integrity and fall apart. Their niche in the theoretical and explanatory scheme of things is reconstructed and reconstrued, and new concepts with quite different dimensions and dynamics come to occupy the newly carved niche. The 'spirits' and 'principles' of alchemy, the 'crystal spheres' of pre-Galilean astronomy, 'daemonic possession' of Medieval medicine, 'phlogiston', 'ether', and 'signatures', are now nought but dry bones of an earlier intellectual ecology. The theme of this paper is that a similar fate may befall concepts respected and revered in our own prevailing conception of how humans works, and the concept on which I mean to focus is consciousness. To the degree that there are already afoot misgivings about the integrity of the traditional conception of consciousness, some of the problems discussed will be familiar, especially to psychologists. But the extent of the erosion of the concept by recent empirical findings is, I suspect, greater than hitherto reckoned. The aspects of the orthodox concept which I shall discuss include the alleged transparency of the mental, the supposed unity of consciousness and the idea of the self, and the allegedly special relation thought to obtain between language and consciousness.

PART I

The venerable dogma that one's mental life is self-intimating and introspectively available seems at last to be ready for consignment to the museum of quaint and antiquated myths about how humans work. To be sure, the psychoanalytic literature had worked some wrinkles into the dogma, but on its own, psychoanalytic theory was perhaps not enough to raise serious and systematic problems.

*I am especially indebted to Donald Griffin, Jack MacIntosh, and Daniel Dennett, who inspired and helped me. This research was supported by the Social Sciences and Humanities Research Council of Canada, grant no. 410-81-0182.

After all, one could always say that it was really only the naughty and nasty bits of one's mental life that sometimes escaped introspective notice, not the exalted and more serious business of cognition and sentience. Moreover, it was generally thought that given the right conditions, even the seamier side of one's mental life could be marched out of the occluding shadows. In other words, it was seen as an aberration from the typical transparency of mental life that certain psychological states and processes escaped even the most discerning introspective eye. Now, however, it is evidence that a great deal of intelligent and sentient activity, perhaps the lion's share, goes on without benefit of self-conscious awareness, and that so far from this being a pathological condition, it is in the nature of the case. Now for the evidence.

On the one hand, there are data showing that the cortex responds to subthreshold stimuli¹; Libet (1967) has shown this for haptic stimuli, Martin (1974) for auditory stimuli, and Lehman (1967) for visual stimuli. On the other hand, there is evidence that stimuli, subthreshold for conscious awareness, are capable of exerting behavioral effects, sometimes of a rather sophisticated kind. Thus has Shevrin (1973) shown the effect of subliminal perception on word choice in subsequent conversation, Kolers (1975) showed the effect on problem solving of prior subliminal presentation of the answers. Lackner and Garrett (1972) have demonstrated related effects using auditory stimuli. Zajonc (1980) has obtained intriguing results indicating that there are exposure effects (i.e. preference for the items one has been exposed to) even when the item was presented for only one millisecond durations, and hence well below the threshold for conscious recognition. Coming at the question from a different perspective, cognitive psychologists have found it theoretically necessary to postulate masses of unconscious processing in order to explain intelligent behavior. Most renowned perhaps of these postulations is owed to Chomsky, who pioneered the argument that overt speech is but the consumation of complicated unconscious processing.

From another but related perspective, social psychologists have come to the conclusion that unconscious cognitive and sentient elements figure in judgments of preference and quality. Thus Nisbett and Wilson (1977) have shown that when subjects are set the task of evaluating candidates for a job by examining the candidates' files, if they are told they will subsequently meet a particular candidate, this typically results in a more favorable judgment of that candidate than if they are not so told. For the evaluators, this effect on their judgment of expecting to meet the candidate was utterly dumbfounding, for they supposed their evaluations to be based entirely on other considerations. (For many other wonderful examples, see Nisbett and Ross, 1980).

In studying the skill of certain blind persons in manoeuvring about obstacles, Supa, Cotzin and Dallenbach (1944) found that echo-location is the crucial mechanism, even though subjects either had no idea of the basis for their skill, or had the fixed idea that they used a complex tactile mechanism exploiting differences in air pressure and sometimes referred to as 'facial vision'. Either way, many of the skilled blind subjects did not attribute a significant role to auditory cues. Nonetheless, in a series of carefully designed experiments it was found that when subjects' ears were stopped up, the manoeuvring and detecting abilities faltered entirely, and that auditory cues were vital. The sound of shoes on a hard floor or the sounds from fingers snapping turn out to be of crucial importance, notwith-

standing introspective reports to the contrary. (For a fascinating discussion of this, see Griffin 1974)

Rather jolting to the transparency thesis is the finding that, unbeknownst to self-consciousness, the size of the pupils of the person one was looking at figured crucially in the judgment that of two pictures of him, one was more friendly, warm and appealing. (Hess, 1975) The jolt in this experiment is delightfully dramatic, for not only was such information concerning pupil size not a conscious consideration in the judging, but additionally, one was not even aware of having taken any notice of pupil size at all. Finally, when the fact that pupil size figured in the judgment is revealed, there is not, for me certainly, the instant rush of recognition, but rather the observer's appreciation of how some part of the animal world, to wit oneself, behaves.

More dramatically still, the discovery of blindsight in humans (Weiskrantz et al. 1974) puts paid to the idea that perceptual judgments require consciously available perceptions. Weiskrantz showed that a subject with lesions in the primary visual cortex which resulted in blindness in his left visual field, could nevertheless make perceptual discriminations about items in the blind visual field with an impressive accuracy. For example, he could point to where, *in his blind visual field* a light was shining, and he could distinguish vertical from horizontal bars of light with an accuracy of 29/30, 30/30. The acuity came as a surprise to the subject, who insisted he was merely guessing his answers because he had no awareness of visual perception in his left visual field. But the accuracy of his discriminations implies that it was no mere guessing, and that nonconscious visual data, perhaps from alternate anatomical pathways, figured in his responses to questions. (For other cases, see Perenin and Jeannerod, 1975)

Even stranger and perhaps more disruptive of our common sense assumptions is the clinical syndrome of blindness denial, also called Anton's syndrome, after the nineteenth century neurologist who wrote on the matter. Certain patients who suffer a sudden onset of blindness due to lesions or trauma to the visual cortex, do not recognize that they have lost their vision and can no longer see. The clinical pattern here is that these patients cannot see, but they are unaware that they are blind, tend to deny their visual deficit, and persist in behaving as though they can see. Often they invent excuses to explain why they bump into the furniture (e.g. "I lost my glasses," "there is too much stuff in here") and they may confabulate² when asked to describe what they see (e.g. when the doctor holds up two fingers and asks the patient how many she sees, she may reply with assurance that she sees four, or she may describe the curtains in the room as red when they are really white, and so on.) Moreover, there appears no convincing reason to suppose the patients are deliberately making things up, nor can their obstinate persistence in believing that they can see quite normally be explained either as a generalized deficiency of the intellect or as a determination to keep a stiff upper lip or as based in hallucination, though in common with many blind patients, they do occasionally hallucinate. The disorder is sometimes permanent, and sometimes transient. Should the patient recover, she typically is not cognizant of her improvement and continues to insist she never had a problem with her vision. Critchley (1979) notes:

Reaction to the disability, we find by experience, varies from simple unawareness of the visual effect to blatant denial of the blindness, reluctant admission that the sight has

perhaps deteriorated to a minor degree, projection of the disability to the environment such as inadequate lighting. (p.119)

In some sense, incredible as it may seem, these patients do not know that they do not see. Whilst it may perhaps be readily imaginable that one could be unaware of the loss of one visual field, it does seem astonishing that one could fail to know that one has no vision at all. Philosophers may be so dismayed and shocked by the idea, that they may insist the syndrome is wrongly described. Several points are relevant here. First, denial of cortical blindness is not rare. Any neurologist in any large hospital will have seen at least one classical case and probably more. Second, what the phenomenological and behavioral and neurological situation is here, is an empirical matter. To be sure, current descriptions by neurologists may be mistaken, but on the other hand, we certainly do not know enough about awareness, about conditions for awareness, about diminution of awareness and so on, to insist that if someone is blind he *must* know that he is blind, whatever his behavior and whatever his neurological status. Normally, we take shutting our eyes as a prototypical way to be not seeing, and since there is all the phenomenological difference in the world between having one's eyes open and having them closed, we suppose that anyone who does not see is in a state like us with our eyes closed and so cannot help but know. However, what the blindness denial syndrome teaches us is that we ought to be prepared to revise our common sense beliefs about awareness. There are, it seems, more ways to fail to see than might have been suspected, and some of these we are not equipped to notice. (For more on this topic, see Hecquen and Albert, 1978; Critchley, 1965)

Finally, to round out the case against transparency and the reliability of introspection, it is pertinent to mention also the homey familiar phenomena noticed outside the laboratory, as when one drives home on 'automatic pilot,' or discovers oneself to have been cranking out proofs for freshmen whilst woolgathering on a distant topic. Of course, having agreed that there must be a tremendous amount of covert intelligent activity does not yet tell us how best to characterize that activity; in particular, it does not tell us that the best or most productive model for understanding mental activity *in general* is the one prevailing in folk psychology. (See my 1980a and 1980b)

Introspection has been found much less thorough and reliable than previously supposed. Confabulation appears to be natural and easy to us, and the naturalness of our self-appraisal and our explanations 'from the inside' can easily be mistaken for privileged access. Gazzaniga (1978) has elegantly demonstrated this confabulatory facility with one of his split brain patients. Each hemisphere was simultaneously presented with a different picture, and the subject was then to select from an array placed before him, the card which best fitted with the flashed picture. In one case, a snowy scene was presented to the right hemisphere, and a chicken's claw to the left. The subject then looked over the cards placed in front of him, and the right hand subsequently chose a chicken's head to go with the chicken and the left hand chose a shovel to go with snowy scene. When asked about the choices, the left hemisphere explained without hesitation: "I saw the claw and I picked the chicken, and then the shovel because you have to clean out the chicken shed with a shovel", thereby explaining his left hand's action. This inventive and appropriate confabulation by the left hemisphere to explain action initiated by the right was observed repeatedly, with no evidence either that the

left hemisphere was consciously mendacious, or that it was puzzled about what the left hand was doing. Luria (1976) also notes the confabulatory tendency in patients with tumors in the medial zone of the frontal lobes. Nisbett and Wilson (1977) have shown this confabulatory facility in ordinary folk. In one study, they placed a number of *identical* pairs of panty-hose on a table in a shopping mall and asked passers-by to rate them for quality. A pronounced position effect was observed, in that the panty-hose in the *right-most* position were typically preferred, though subjects explained their choice as owed to differences in quality, such as superior strength, sheerness, color, and what have you, which they thought they had observed. (See also Goethals and Recheman 1973, and Storms and Nisbett 1970) There is no evidence that such confabulation is pathological; rather it appears to be a normal part of theorizing about persons and their behavior, oneself and one's own behavior included.

If the elements of conscious deliberation may sometimes play second fiddle to conscious causes of deliberate action, if unconscious states can figure in choice and judgment, then the old dogma is beleaguered also in its insistence that consciousness and control go hand in glove. Control may turn out to be only incidentally and occasionally connected with awareness, and the dynamics of human behavior may turn out to bear but passing resemblance to the prevailing notion of the conscious self in control of cognitive activity and deliberate behavior. The widespread *feeling* of conscious control may be part of our confabulatory practice, reflecting hapless but guileless error in folk psychology. Moved by their studies of split brain patients, Gazzaniga and LeDoux (1978) are intrigued by the hypothesis that the mind is not a unified psychological entity, but a 'sociological' entity, in the sense that *control is distributed*, not deposited despotically in self-consciousness. It is part of their view that the conception of the unified self it at bottom illusory, the product of theorizing, where some of that theory is presumably encephalized as one learns the language. Whether they are right is an empirical question.

When Hilgard (1977) argues that the unity of consciousness is an illusion, he points not only to the rather remarkable dissociation phenomena which characterize fugues and split personalities. Doing several things at once, so commonplace and prosaic as to pass almost unremarked, also fails to square with the idea of unity of control. It is to be marvelled that one can follow a conversation whilst preparing a reply, balancing one's glass and plate, and keeping an eye out for the host; one can read children a bedtime story whilst intermittently musing on how blood clots. In speaking, a person standardly not only attends to what he is saying, but also monitors the non-verbal responses of the listener, both of which feats are highly complex. Ostensibly, control is divided, though to be sure, the dynamics of that, or what it really comes to, remain profoundly baffling. Must there be a supervenient controller who divides subsidiary control, or is it possible that a strict hierarchical model of control is wrong, that control shifts and moves in a quite different dynamic? This is, finally, an empirical question, a question about how in fact the brain works, and we should be ready to find that a strictly hierarchical model of control may be as inappropriate to the brain's organization as the model of intelligent creator was to the organization of Nature.

From what we do know of the brain, it seems highly implausible that there is a pontifical neuron, or a pontifical-neuron group which acts as the control centre.

In wondering how control might be managed by the brain, it is useful to reflect that in computer programming, control can take several forms; e.g. it can be given to a central executive who metes out direction and chores, or it can take a 'bulletin board' format, in which case problems are posted or 'broadcast', and the appropriate unit takes control and relinquishes it when the job is done. Does the brain perhaps conform to this second mode of operation? It is certainly too early to tell, but what is useful about the second style is that it provides an alternative to the executive control model, and in so doing, frees us from the idea that executive control is essential and inevitable, and permits us to wonder about models different yet again, which might be helpful in figuring out how the brain manages. It may be that in the brain a variety of factors have to be taken into account, such as a genetically set rhythm for shifting control (e.g. for waking, REM sleep, non-REM sleep) or for left hemisphere-right hemisphere dominance (Klein and Armitage 1979). Other factors might include learned cognitive habit and physiological relevance of in-coming signal.

If our conception of control changes, then our conception of the self change *pari passu*, and this possibility may initially produce a kind of dizziness, tainted perhaps with anxiety. The warm antiquity of self, as Wallace Stevens calls it, binds us profoundly. Even so, a new theory of what we are could well be immensely liberating and enlightening, and a reconfiguration of the concepts might well provide us with new possibilities for understanding the mind/brain which were hitherto undreamt of. The self, in the event, may turn out to be not so much a unified coherent entity, but a disconnected collection of wants, needs, and what-nots, flying in loose formation.

PART II

The role of linguistic representation in all this deserves some pondering, for it is a common enough theme that there is a special relation between consciousness and linguistic competence. Different as they are in metaphysical beliefs, Gazzaniga and Eccles nevertheless find common accord in the idea that human consciousness is what it is because of verbal competence. Gazzaniga and LeDoux (1978) suggest that the difference between conscious and unconscious states emerges only with the acquisition of verbal facility. Thus they say that "...the verbal system's role in creating a sense of conscious reality is critical and enormous." (p. 155) Like Eccles (1977), they deny the right hemisphere the conscious states attributable to the left, dextral and impressive manipulo-spatial performance of the right hemisphere notwithstanding. Philosophers as thoroughly contrapuntal in their views as Malcolm and Dennett likewise voice in concert the special relation between language and consciousness. The idea is widespread, and may perhaps be granted the status of conventional wisdom.

Its chartered respectability notwithstanding, this view asserting the special relation between language and consciousness deserves almost as much suspicion as the companion supposition concerning the transparency of mental life. In what follows, I want to wind up a number of suspicions and set them afoot for observation and consideration. In so doing, I certainly shall not claim to have proved the falsity of the thesis in question, and indeed I should be wary of any *a priori* argument to that effect. Rather, I shall argue that it is an empirical question

whether language is essential for consciousness, and that the empirical evidence available so far fails to confirm the hypothesis that there is a special relation.

It is appropriate to begin by considering how the relation between language and consciousness is specified by those who hold it to be special. The astonishing thing is that substantive answers here are exceedingly rare, both in the work of hands-on empirical researchers, and in the work of those who research is removed from the laboratory. It is almost as though it were too obvious to merit explanation. For example, from Gazzaniga and LeDoux, virtually nothing is offered by way of explanation. They say merely that language *creates* consciousness, or that it makes consciousness possible. However, such remarks do not find their way to the explanatory side of the ledger; rather, they articulate what needs to be explained. Eccles believes that the non-physical mind makes contact with the brain essentially through the areas of the brain involved with language, for example Broca's area and areas 39 and 40. Popper argues that consciousness is an emergent property and language is somehow the vehicle of its emergence. Tantalizing as these ideas may be, they are in no case an explanation of what it is about consciousness that requires language, nor what it is about linguistic representations that somehow causes, creates, or otherwise brings about consciousness.

The matter is worth pressing because of course it is not at all obvious that there is a special relation between language and consciousness. Indeed, as some ethologists are wont to point out, it is starkly evident that the higher animals at least, *are* conscious, and that these canny but dumb creatures manage to pull this off despite lacking membership in a human language community. Additionally, preverbal children and certain nonverbal adults, such as patients with left hemisphere lesions, or deaf mutes, certainly seem to be conscious. There are therefore, powerful *prima facie* reasons for doubting that language is essential for conscious experience, and in consequence for wondering whether the significance of language for cognitive functions has been misestimated. (See my 1978, 1980a). The hypothesis that language is crucial for consciousness accordingly needs to be investigated, for the idea might be funded by nothing less inglorious than mere prejudice in favor of one's species.

An argument which has appealed to some philosophers is that there is a conceptual connection between consciousness and language. The basic idea here is that thoughts are necessarily linguistic, and the contents of consciousness are necessarily thoughts. Now the 'necessity' which does all the work in this argument is supposed to derive from facts about what is meant by certain expressions. That is, there are conceptual relations connecting thought and language, and connecting consciousness and thought such that we allegedly mean something linguistic in character by 'thought' and something characterized by thought when we talk of consciousness. So, for example, Vendler (1972) says:

If, following the hint of etymology (con-scire), consciousness be defined as the totality of one's thought (at a given moment), then sensations and feeling do not enter one's consciousness until by noticing or being aware of them one forms or entertains a perceptual judgment. If this is true, then we must agree with Descartes that animals, strictly speaking, cannot be conscious of their sensations and other experiences. (p.162-3)

(For a similar account, see Hartnack, 1972) This is unsatisfying. For one thing, it is not clear that the described connections are 'conceptual', or whether this is an instance of a philosopher's conjuring the conceptual in aid of a favored but faltering hypothesis. A more serious reservation is this: even if the meaning of 'consciousness' is tied to the meaning of 'linguistic representation', such a fact of language does not determine the nature of things; that is, it does not determine whether consciousness really does require language, and why, or whether animals, because they lack a human language, cannot be conscious in something like the way humans are conscious. If we are entertaining questions about the psychological states of humans and other animals, then it avails us nothing to find a definitional tie between consciousness and language use, and then assert, with the air of having discovered something, that animals are not conscious. Special definitional relations can anyhow be made to succumb when the empirical pressures demand it. A special definitional relation between physical space and Euclidean geometry did nothing to save Euclidean space from Einstein's theory and Eddington's observations.

The uncompromising Cartesian view that animals are mere automata, incapable even of feeling pain, seeing red or hearing sounds, is nowadays regarded as impossible to defend without appearing downright ludicrous. There is, however, a compromise strategy which both conserves the Cartesian idea that human consciousness is in a class by itself, and at the same time throws something to the clamouring ethologists. The strategy is this: draw a distinction which will accord a lower, brutish grade of consciousness to nonverbal animals, and will accord a high grade or full-blooded consciousness to humans. This serves to gratify the conviction that dogs do indeed feel something unpleasant when kicked, but at the same time preserves the special status for human awareness. Language is of course what makes the difference between these two (natural) kinds of psychological state. Thus Vendler, having said that animals are not strictly speaking conscious, goes on to relax his strict speaking by advancing the hypothesis that animals are aware of "raw feels", but they have no states akin to propositional attitudes — nothing analogous to beliefs, desires, intentions, and the like. Gazzaniga and LeDoux (1978) are similarly persuaded, and whilst allowing that preverbal children, animals and the right hemisphere have a *sort* of consciousness, they insist that it is different in *kind* because it is nonlinguistic. A variation on this theme is that the nonverbs are merely conscious in contrast to the verbals, who have *self-consciousness*. In this spirit, Dennett (forthcoming) says:

I suggest that we should entertain the hypothesis that what we had taken to be one phenomenon is actually two quite different phenomena: the sort of consciousness that is intrinsically connected to the capacity to say in one's natural language what is going on; and the sort of consciousness which is just a matter of intelligent processing. On this proposal, adding the capacity to make "introspective reports" *changes the phenomenon*, so when we are wondering what a dolphin or a dog could tell us, or what a non-dominant hemisphere could tell us, if only they could talk, we are wondering about a radically different phenomenon from the phenomenon that exists in the absence such a linguistic capacity. (p.12)

If I understand him aright, the suggestion seems to be that for the nonverbal, conscious life is rather like my nonconscious mental life — like the part that manipulates semantic and syntactic structures before I speak, or the part which ‘sees’ in subliminal perception — all hidden from awareness, but for all that, purposeful and efficient.

What troubles me about settling into the ecumenical cosiness provided by the distinction between natural kinds of consciousness, is that it is not at all evident that the distinction carves nature at her legendary joints, or whether it serves only to shelter intuitions which need to be schooled. For one thing, it appear as though there is not just one distinction, but a set, joined by the common purpose of finding human consciousness unique in the animal kingdom. For another, all the earlier concerns with specifying the difference language makes can simply be restated. What is it about self-consciousness such that it requires linguistic representations, and what is it about language such that it brings about the special capacity for self-consciousness? What are the empirical grounds for supposing animals are deprived of this capacity? *How do we know that adding the capacity to make introspective reports radically changes the phenomena?*

If there was a paucity of explanation earlier of the relation between consciousness and language, that lack is not made good with respect to self-consciousness and language. Still, a general motivating hypothesis can be discerned, and it is roughly this: an organism has representations if and only if it has a linguistic medium for those representations, and an organism is self-conscious only if it has representations. The problem with the argument is this: if by “language” is meant “human-like language”, then it is highly implausible that an organism must have a language in order to have representations. The complexity of behavior in many species of animals invites, nay *requires*, the postulation of an internal system of representations. The artful and complex orienting and communicative behavior suggests that nonverbal organisms have a cognitive map of the environment, preference and decision structures and learning routines, and that they have the representational wherewithal to permit them complex interaction with their own and sometimes other species, where such interaction often includes a division of labor and a recognition of individuals as individuals. Watching dogs herd sheep, or chimpanzees fashion dipping sticks for termite foraging is salutary in this respect. Certainly it is puzzling to deny that preverbal children lack a representational organization, for such children routinely and easily engage in remarkably sophisticated cognitive business, not the least of which is learning language. If the preverbal child is representationally bereft, language learning is sheer magic.³

In both preverbal children and animals, there appears to be cognitive activity which may be thought of as the nonverbal analogue of overtly linguistic theorizing. Insofar as there is, their ‘inner life’ can be expected to have a structure and dynamic which yields a picture of the world and of themselves in it. Human adults with Broca’s aphasia or with left hemispherectomies are also appropriately thought of as capable of cognitive activity of a nonlinguistic variety. We may not yet know how to characterize representational states save via our folk psychology which takes propositional attitudes as the model for representational states generally, but this should properly be seen as a limitation on folk psychology, rather than a limitation on what kind of creatures have representations. It is certainly a

mistake to assume that this model provides an essentially correct model of representations generally, but even if it does, notice that the Dennett hypothesis is enfeebled. For if the representational capacity of nonverbal animals be explained in terms of a language of thought, then there is no reason to suppose that acquiring the capacity for *overt* linguistic behavior 'radically changes the phenomenon'.

Accepting our evolutionary bonds to nonverbal animals, we should be prepared to find that they do indeed have brains which conduct intelligent business, which learn, and which enjoy consciousness. After all, as Donald Griffin (1981) remarks, if we can find instructive similarities between hominids and other animals with respect to bone structure, embryological development, antibodies and the nervous system, then it is to be expected that we will find them also with respect to cognition and consciousness. Only the most slavish verificationist would gainsay self-consciousness in nonverbal animals on grounds that verbal behavior is, at this stage of our knowledge, the only sign of self-consciousness which is universally accepted.

The fact that we use reportability on the part of a verbally competent human as a test for whether or not the stimulus was subthreshold (p. 2 and note 1) does not imply that the same test is appropriate for the nonverbal human or other nonverbal animals. Reportability is in any case, a *provisional* criterion, usable, warts and all, until better physiological measures are forthcoming from a more detailed neuroscience. Criteria for determining when something belongs to a certain kind evolve as theory evolves, and the originating criteria may even be set aside if the advance in theory involves a retaxonomizing of kinds. By way of illustration, consider something as basic as criteria for gender. Before molecular biology, the determination of the sex of an animal was based decisively on inspection of genitalia; male genitalia meant male animal. Now we know that things are more complicated, and that a *genetic* female (XX chromosome pair) can have male genitalia, if certain other parts of the genetic package are not in order. Additionally, as the effects of sex hormones on the developing brain are beginning to be understood, it is evident that there can be a mismatch between brain gender and genetic gender, and between brain gender and genitalia gender. The point of the analogy, therefore, is that reportability as a test for awareness in verbal humans cannot be assumed to be immutable or to be extendable to nonhumans or to nonverbal humans, even though the test is now useful in determining that non-conscious states affect human behavior.

Even if language is not necessary for representation, is it nevertheless required for self-awareness, and may it be that with the acquisition of language there does indeed emerge a new psychological phenomenon? This seems to be the hunch fuelling Dennett's idea that one sort of consciousness is intrinsically connected with the capacity to introspectively report. Now suppose we can say, in a rough and ready way, that what is required for self-consciousness is that there is enough richness in the representational system such that the organism goes into representational states about itself on a noninferential basis. Language comes in this way: an organism would need to entertain reflexive sentences in some natural language in order to be able to do this. That is, it would need to be able to say such things as "I am grooming me" in order to have self-grooming representations.

Why should we be convinced of this? Trying in one's own case to have self-representations without using English (or what have you) may indeed provoke intuitions supporting Dennett's hypothesis, but unfortunately the test does not show very much. It may be a bit like trying not to see an illusory contour after having had it come into focus, since for fully verbal humans, some part of the self-representing undoubtedly is linguistic. However, this does not entail either that it all is, or that language is essential to self-representing. If an organism can enjoy a complex and sophisticated representational system in virtue of which it hunts and forages, and gets on in its social community, communicates with conspecifics, identifies others in his group as individuals — *all without linguistic representations* — there is nothing very remarkable in supposing it has representational states about itself as well. The possibility that a monkey, say, could have non-linguistic representational states functionally analogous to reflexive relations, such as 'x hit x' and 'x grooms x', seems no harder to imagine than the possibility already granted, that it has a cognitive map, that it has the non-linguistic analogue of a theory of social relations in its community, and that it learns via a non-linguistic analogue of hypothesis formation and testing. Indeed, it may have all these in the way most hominids for most of the species' history have had them. Seen in this light, one would expect that many representational systems would have the wherewithal to permit some degree of self-consideration. It may be that an organism cannot have high-falutin self-representation without a language, so that a chimpanzee might never ruminate on the evils of self-indulgence, or experience existential torment, but such deprivation he undoubtedly shares with many a fully verbal human. In any case, it is insufficient to distinguish his consciousness as different *in kind* from the human sort. For all we can see at this stage of our knowledge, reflexive representations, such as 'x grooms x' seem not to be marked by a difference in kind but a difference in degree of sophistication of the representational system. My point then is an essentially negative one: if self-representations are to be classed as special in that they are conditional upon a linguistic capacity, then special reasons need to be advanced for this classification, for why increased complexity of non-linguistic resources of a representational system could not suffice. (See also Sellars, 1980)

Striking data concerning self-representation in a nonverbal species was ingeniously collected by Gordon Gallup (1977). Gallup's strategy was to test chimpanzees for self-recognition in mirrors. In his set-up, a number of wild-born, pre-adolescent chimpanzees were separately put into rooms empty but for mirrors. Initially a chimpanzee responded to his mirror image as though it were a stranger, just as infant humans initially do. Within two or three days, however, the chimpanzees began to use the mirror for self-directed behavior, such as grooming hard-to-see spots, picking the teeth, and making faces. Such behavior renders plausible the hypothesis that the chimpanzee recognizes himself in the mirror.

In order to test the hypothesis more thoroughly, Gallup devised a further experiment. After having anaesthetized the animals, he put a dab of red paint (odorless and non-irritating) over one eyebrow of the chimpanzees, and after recovery they were permitted, one by one, to view themselves in the mirror. In front of the mirror the chimpanzee fingered the spot gingerly and frequently (these 'mark-directed' actions were counted and compared to the count of touches made elsewhere), viewed themselves studiously, and in a number of cases the chimpanzees

examined their fingers after having rubbed the strange spot. Chimpanzees innocent of mirrors were similarly dabbed under anaesthetic and showed none of the self-examining behavior of the experimental animals.

Gallup concludes that chimpanzees do indeed have a self-concept, in virtue of which they readily learn to recognize their own reflections. He discusses the fact that self-recognition in mirrors seems beyond many other species, including rhesus monkeys, macaques, java monkeys, and baboons, despite almost heroic efforts on the part of the experimenters. (Some success has been achieved with orangutans.) A number of reasons could explain this inability, and it would be rash to conclude that failure implies the absence of a self-concept, though such absence is consistent with the data. One possibility is that the animals' self-concept is insufficiently rich or adaptable to result in success at self-recognition in a mirror. This idea implies that concepts of the self may come in varying degrees of sophistication. Thus Gallup (1977) suggests:

Different organisms may very well have differing degrees of self-awareness, but only with an explicit sense of identity does self-recognition become possible. The threshold for self-recognition [in mirrors] may be quite high compared to other forms of self-conceptions. It is possible that as yet unidentified tests of the self-concept, which require a lower threshold for recognition, might yield positive evidence for monkeys. (p. 334)

Gallup is not claiming that this is indeed so, but only that it is an empirical possibility consistent with the available data, and well worth pondering. It is also noteworthy that some profoundly retarded humans seem incapable of recognizing themselves in mirrors (Gallup 1977 p. 331) though Down's syndrome children typically can. (My observation)

The suggestion that the self-recognition test with mirrors may in fact be an imperspicuous test for the presence of a self-concept is intriguing, for organisms did not evolve in an environment replete with mirrors, and accustomed as we are to mirrors, we tend to forget how very, very odd they are. In animals less cognitively nimble than chimpanzees, the plasticity and sophistication to handle such novelty may be lacking, even if the creature had a primitive form of self-conception. It is easy to imagine aliens who out-strip us intellectually, and who decide to test us for self-awareness by using a device for self-recognition which is familiar to them, but which requires information processing at a level of complexity quite beyond us. When we fail their test miserably, we would bemoan their failure to find the appropriate test where we could display great prowess in self-awareness.

Interaction with conspecifics may be important for the proper development of a self concept, for it may be that in order to have a conception or model of the kind of thing one is, one needs models of others. Barlow's (1979) thoughts on this question seem insightful:

Self-awareness would then result from a brain modelling the reaction of other brains, and incorporating the fact that the others, like itself, are nodes in an interacting network. This recognition that others are unique but like oneself, implies the reciprocal, which is self-awareness: "I am unique, but similar to others." Self-awareness is a product of the efficient modelling of the relations between brains. (p.82)

The brain apparently needs to interact with the environment in order to generate a body image, and it may be that it needs to interact with other brains in order to generate its 'brain image'/model of itself, though such a model would be opaque, in the sense that it would not contain information that it is a model of the brain. An unusual feature of Barlow's description is that it does not specify that the interaction must be verbal; rather he seems to recognize that verbal interaction is only one kind amongst many others which will serve the brain in generating its model of other brains. It would be needlessly narrow to adopt the theory that the brain models by singling out amongst behavior the overt utterances it hears other brains emit, and sees in the structures and relations it therein discovers, the constituting pattern of how other brains work. How brains, human and otherwise, model the world, model other brains, and finally themselves, is truly a mystery of mysteries. What we should avoid in trying to solve the mystery is the idea that linguistic representation is the prototype for representation generally. Accordingly, we should seek models of representation which operate on radically different principles. (For a suggestive attempt, see Edelman 1979)

Language may be important to this extent: some of the theories we learn are learned via a language, and as creatures differ in theories which structure their epistemic lives, so they differ in their mental lives. However, that does not make language the be-all, nor does it serve to mark the difference in kind for which many thinkers variously plump. It is worth wondering whether such differences as do obtain might not simply be on the order and significance as those obtaining between verbally accomplished organisms with very different theories about the world. My point is that for all we know, the differences between my conscious states and those of a gorilla may be no more differences in *kind*, and no more significant, than those obtaining between a medieval serf and a modern man steeped in biological and physical theory, or between a child with Down's syndrome and a professional musician, between Captain Fitzroy and the inhabitants of Tierra del Fuego, between oneself at five and at sixty-five.

With the rout of a dogma, much that seemed obvious comes to seem deeply puzzling, and thus it is with the nature and function of consciousness. Thus freed, the imagination may wonder whether, from the point of view of the brain, the difference between states which are conscious and those which are not, is in fact very significant. For all our nescient conviction, it remains an empirical possibility that the difference is rather minor. A more fundamental and significant difference might divide representational from non-representational states, where the latter might be yoked to essentially housekeeping duty. Such a distinction could be expected to cut across the conscious-nonconscious boundary. In our naivete, it seems now that conscious states are a single, unified, *natural kind* of brain state, but we should consider the possibility that the brain is fitted out with a battery of monitoring systems, with varying ranges of activity and with varying degrees of efficiency, where consciousness may be but one amongst others, or where these systems cross-classify what we now think of as conscious states. States we now group together as conscious states may no more constitute a natural kind than does say, dirt, or gems, or things-that-go-bump-in-the-night.

Finally it must be asked, how should we try to understand what consciousness is? A variety of research strategies and approaches may be fruitful, but it seems to me that they should finally converge upon, in one way or another, investigation

of the brain itself. In addition to research already discussed, studies to determine the neurophysiological differences between conscious and nonconscious states, to find out what goes on in the brain during REM sleep and during non-REM sleep, to determine the order and nature of events during the various facets of attention, and how this fits into a broader scheme of intelligent brain activity, would surely contribute to a deeper understanding of what sort of business consciousness is. Examples of current research which seem to me to be very exciting would include Mountcastle's single cell recordings (1978) in the studies of visual attention, which provide a gripping glimpse of the detailed neuronal organization subserving various facets of visual attention. The carefully detailed study of how brain stem structures figure in arousal and alertness (Jouvet 1967; Scheibel and Scheibel, 1967; McCarley 1980; Lindsley and Wilson 1975), and of the role of hippocampal structures in short term memory and attention (Vinogradova 1975) are also important. From yet another direction, the work done on functional localization by determining variations in uptake of oxygen and glycogen in distinct areas of the brain whilst it is engaged in a variety of tasks, looks to be an extraordinary and powerful development. (See Lassen, Ingvar, and Skinhoj, 1978) Ethological studies (see, e.g. Kummer 1982; Seyfarth, Cheney and Marler, 1980) are also of great importance here for a variety of obvious reasons, but also because they help us to see that mere brains, unaided by gods, souls, freely exercising wills and human language, can nonetheless exhibit stupendous complexity. The hope is that the assorted types of investigation will inform and inspire one another, as we struggle free of confining idiom, and come to understand what manner of creature we are.

Institute for Advanced Study
Princeton, New Jersey
and
University of Manitoba
Winnipeg, Manitoba

NOTES

¹The stimulus is subthreshold for awareness in a verbally competent subject if he cannot report the occurrence of the stimulus and denies that he perceived it. (See also p. 10 below) For determining threshold levels, see Shevrin (1973) and Dixon (1971).

²In the neurological and psychological literature, 'to confabulate' is used to mean 'to invent', 'to rationalize or make thing up', and it carries the connotation that the rationalizing and inventing is not like playing make-believe or intentionally fibbing; rather, the person believes his confabulations. To be sure, the standard dictionary meaning is quite other, to wit: 'to talk together'. How 'to confabulate' came to acquire its application in neurology and psychology is a mystery to me though it suggests a running together of 'to contrive' and 'to fabulate'.

³Dennett is clearly aware that there is a problem with these cases, for he says:

"How should we conceive of the gradual acquisition of language by children (or chimpanzees, perhaps!), and what should we say about the "experience" (conscious or unconscious) of the pre-verbal Helen Keller? A theory of consciousness that does justice to the variety of complications will almost certainly demand of us a revolution in our habits of thought about these demand issues, but we have already absorbed the first shocks of that revolution in the overthrow of the Lockean vision of the transparent mind." (p.13)

REFERENCES

- Barlow, H. B. (1979) "Nature's joke: A conjecture on the biological role of consciousness." In: *Consciousness and The Physical World*. Edited by Josephson and Ramachandran (Permagon:London), 79-92.
- Churchland, Patricia Smith (1978) "Fodor on language learning", *Synthese*, 38, 149-159.
- (1980a) "Language, thought, and information processing", *Nous*, 14, 147-170.
- (1980b) "A perspective on mind-brain research", *The Journal of Philosophy*, 77, 4, 185-207.
- Critchley, Macdonald (1965) "Modes of reaction to central blindness", *Cesk. Neurology*, 28.
- (1979) *The Divine Banquet Of The Brain*, (Raven:New York).
- Dennett, Daniel C. (forthcoming) "Consciousness". In: *Oxford Companion To The Mind*. Edited by Richard Gregory, (Oxford University Press: Oxford).
- Dixon, N. F. (1971) *Subliminal Perception: The Nature Of A Controversy*, (McGraw-Hill:London).
- Eccles, J. C. and Popper, Karl (1977) *The Self And Its Brain*, (Springer-Verlag:Berlin).
- Edelman, Gerald M. (1979) "Group selection and phasic reentrant signalling: A theory of higher brain function." In: *The Neurosciences Fourth Study Program*. Edited by Francis O. Schmitt and Frederic G. Worden. (MIT Press: Cambridge, Mass.), 1115-1139.
- Gallup, Gordon G. (1977) "Self-recognition in primates: A comparative approach to the bidirectional properties of consciousness", *American Psychologist*, 329-338.
- Gazzaniga, Michael S. and LeDoux, Joseph E. (1978) *The Integrated Mind*, (Plenum:New York).
- Goethals, G. R. and Rechman, R. F. (1973) "The perception of consistency in attitudes", *Journal of Experimental Social Psychology*, 9, 491-501.
- Griffin, Donald R. (1974) *Listening In The Dark*, second edition, (Dover:New York).
- (1981) *The Question Of Animals Awareness*, second edition, (Rockefeller University Press:New York).
- Hartnac, Justus (1972) "On thinking", *Mind*.
- Hecaen, Henry and Albert, Martin L. (1978) *Human Neuropsychology*, (Wiley:New York).
- Hess, Richard K. (1975) "The role of pupil size in communication", *Scientific American*, 233, 5 110-118.
- Hilgard, Ernest R. (1977) *Divided Consciousness: Multiple Controls In Human Thought And Action* (Wiley:New York).
- Jouvet, M. (1967) "Neurophysiology of the states of sleep", *Physiological Review*, 47, 117-177.
- Klein, R. and Armitage R. (1979) "Rhythms in human performance: 1 1/2 hour oscillations in cognitive style", *Science*, 204, 4399, 1326-1327.
- Kolers, Paul (1957) "Subliminal stimulation in problem solving", *American Journal of Psychology*, 70, 437-442.
- Kummer, Hans (1982) "Social knowledge in free-ranging primates". In: *Animal Mind-Human Mind*. Edited by Donald R. Griffin, (Springer-Verlag:Berlin)
- Lackner, J. R. and Garrett, M. (1972) "Resolving ambiguity: effects of biasing context in the unattended ear", *Cognition*, 1, 359-72.
- Lassen, Niels A. Ingar, David H. and Skinhoj, Eric (1978) "Brain function and blood flow", *Scientific American*, 62-71.
- Lehman, D., Beeler, G., and Fender D. (1967) "EEG responses to light flashes during the observation of stabilized and normal retinal images", *Electroencephalography and Clinical Neurophysiology*, 22, 135-42.
- Libet, Benjamin, Alberts, W. W., Wright, E. W. Jr. and Feinstein, B. (1967) "Responses of human sensory cortex to stimuli below threshold for conscious sensation", *Science* 158, 1597-1600.
- Lindsley, Donald B. and Wilson, Charles L. "Brainstem-hypothalamic systems influencing hippocampal activity and behavior". In: *The Hippocampus—Volume 2: Neurophysiology and Behavior*. Edited by Robert L. Isaacson and Karl H. Pribram. (Plenum: New York).
- Luria, A. R. (1976) "The human brain and conscious activity". In: *Consciousness and Self-Regulation*. Edited by G. E. Schwartz and D. Shapiro (Plenum:New York)
- Malcolm, Norman (1973) "Thoughtless Brutes", *Proceedings And Addresses Of The American Philosophical Association* Vol. 46, 5-20.
- Martin, David G., Hawryluk, Garry A., and Guse, Linda L. (1974) "Experimental study of unconscious influences: ultrasound as a stimulus", *Journal of Abnormal Psychology*, 38, 6, 589-608.

McCarley, Robert W. (1980) "Reticular and aminergic neurons possibly mediating sleep state control", *Neurosciences Research Program Bulletin*, 18,1, 101-112.

Mountcastle, Vernon B. (1978) "Brain mechanisms for directed attention", *Journal of the Royal Society of Medicine*, 71, 14-28.

Nisbett, Richard, and Ross, Lee (1980) *Human Inference: Strategies and Shortcomings Of Social Judgment* (Prentice-Hall:Englewood Cliffs, N. J.)

——— and Wilson, Timothy deCamp (1977) "Telling more than we can know: verbal reports on mental processes", *Psychological Review*, 84, 3, 321-259.

Perenin, M. T. and Jeannerod, M. (1975) "Residual vision in cortically blind hemifields", *Neuropsychologia*, 13, 1-7.

Scheibel, M. E. and Scheibel, A. B. (1967) "Anatomical basis of attention mechanisms in vertebrate brains". In: *The Neurosciences: A Study Program*. Edited by G. C. Quarton, T. Melnechuck and F. O. Schmitt. (Rockefeller University Press: New York), 577-602.

Sellars, Wilfrid (1980), "Behaviorism, language, and meaning", *Pacific Philosophical Quarterly*, 61, 1&2, 3-25.

Seyfarth, Robert M., Cheney, Dorothy, and Marler, Peter, (1980) "Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication." *Science*, 210, 801-3.

Shevrin, Howard (1973) "Brain wave correlates of subliminal stimulation, unconscious attention, primary- and secondary-process thinking, and repressiveness", *Psychological Issues*, 8, 2, Monograph 30, 56-87.

Vendler, Zeno (1972) *Res Cogitans* (Cornell University Press: Ithaca, N. Y.)

Weiskrantz, L., Warrington E. K., and Saunders, M. D. (1974) "Visual capacity in the hemianopic field following a restricted occipital ablation", *Brain*, 97, 709-728.

Zajonc, R. B. (1980) "Feeling and thinking: Preferences need no inferences", *American Psychologist*, 35, 2, 151-175.