

# Consciousness, Unconsciousness and Intentionality

John R. Searle\*

One of the most amazing things about the past half century or so in analytic philosophy of mind is the scarcity of serious work on the nature of consciousness. Even works purportedly about consciousness have very little to say about the general structure of consciousness or about its special features. Thus for example of three recent books containing "Consciousness"<sup>1</sup> in their titles not one contains even an attempt to state a comprehensive account of the structure of conscious states, much less state a general theory that will account for subjectivity, the stream of consciousness, the qualitative character of conscious states, etc. In each case consciousness is regarded not as a primary subject matter in the philosophy of mind but as a "problem", a potential embarrassment to the author's theory, which is,

---

\*© John R. Searle.

<sup>1</sup>Churchland, Paul M. (1984), Jackendoff, Ray (1987) and Lycan, William G. (1987).

in each case, some version of “functionalism” or “materialism” or “computationalism”. What goes for the philosophy of mind also goes for most —not all— mainstream philosophy of language. You can test this for yourself: Which of the many works on meaning by, say Carnap or Quine, has given you the greatest insight into the special features of the relationships between meaning and consciousness?

I think there are many deep reasons for this fear of consciousness in contemporary analytic philosophy. One of the most important of these is that the presuppositions and methods of contemporary philosophy are ill equipped to deal with the *subjectivity* of conscious states. A similar reluctance to accept the consequences of ontological subjectivity also afflicts psychology and cognitive science generally.

In order to account for the mind without consciousness, one must postulate some other sorts of phenomena. The most obvious solution is to postulate unconscious mental processes to do the jobs that traditional philosophy, psychology, and common sense ascribed to conscious mental processes. This manoeuvre takes different forms, but the general tendency in cognitive science has been to drive a wedge between, on the one hand, conscious, subjective mental processes, which are not regarded as a proper subject of scientific investigation; and, on the other hand, those which are regarded as the genuine subject matter of cognitive science, and which, therefore, must be objective. The general theme is always that the unconscious mental processes are more significant than the conscious ones. Perhaps the earliest canonical statement is in Lashley’s claim,<sup>2</sup> “*No activity of mind is ever conscious*” (Lashley’s italics). Another extreme version of this approach is to be found in Jackendoff’s claim<sup>3</sup> that in fact there are two “notions of mind”, the “computational mind” and the “phenomenological mind”. I believe that it is a profound mistake

---

<sup>2</sup>Lashley, Karl (1956). I don’t think Lashley means this literally. I think he means that the processes by which the various features of conscious states are produced are never conscious. But even that is an overstatement; and the fact that he resorts to this sort of hyperbole is revealing of the theme I am trying to identify.

<sup>3</sup>Jackendoff, Ray (1987).

to try to describe and explain mental phenomena without reference to consciousness. In this article I will argue that any intentional state is either actually or potentially a conscious intentional state, and for that reason cognitive science cannot avoid studying consciousness. The attribution of any intentional phenomena to a system, whether "computational" or otherwise, is dependent on a prior acceptance of our ordinary notion of the mind, the conscious "phenomenological" mind.

In order to substantiate this claim, I am also going to have to explore the notion of an unconscious mental state. The course of this exploration is rather more complex than I would like, and I need to begin by reminding the reader of the distinctions between ontology, causation, and epistemology. For any phenomenon, but for biological phenomena especially, we need to know:

1. What is its mode of existence? (ontology)
2. What does it do? (causation)
3. How do we find out about it ? (epistemology)

So, for example, if we were examining the heart, the answer to our three questions is: the heart is a large piece of muscle tissue located in the chest cavity (ontology); the heart functions to pump blood throughout the body (causation); and we find out about the heart indirectly through such methods as using stethoscopes, cardiograms and taking pulse, and directly by opening up the chest cavity and looking at the heart (epistemology). Now, these distinctions apply to both conscious and unconscious mental states. The history of the philosophy of mind in recent decades is in large part a series of confusions between these three questions. Thus, behaviorism confuses the epistemology of the mental with the ontology (we find out about mental states by observing behavior, so mental states just consist in behavior and dispositions to behavior). And functionalism consists in confusing the causation with the ontology (mental states have causal relations to input stimuli, other mental states, and output behavior, so mental states just consist in having these causal relations).

Our naive, pre-theoretical notion of an *unconscious* mental state is the idea of a conscious mental state minus the consciousness. But what exactly does that mean? How could

we subtract the consciousness from a mental state and still have a *mental* state left over? Since Freud, we have grown so used to talking about unconscious mental states that we have lost sight of the fact that the answer to this question is by no means obvious. Yet it is clear that we do think of the unconscious on the model of the conscious. Our idea of an unconscious state is the idea of a mental state that just happens then and there to be unconscious; but we still understand it on the model of a conscious state in the sense that we think of it as being just like a conscious state and as one which in some sense could have been conscious. This is clearly true, for example, in Freud, whose notions of both what he calls “preconscious” and “unconscious” states are built on a rather simple model of conscious states.<sup>4</sup> Perhaps at its most naive, our picture is something like this: unconscious mental states in the mind are like fish deep in the sea. The fish that we can’t see underneath the surface have exactly the same shape they have when they surface. The fish don’t lose their shapes by going under water. Another simile: unconscious mental states are like objects stored in the dark attic of the mind. These objects have their shapes all along, even when you can’t see them. We are tempted to smile at these simple models, but I think something like these pictures underlies our conception of unconscious mental states; and it is important to try to see what is right and what wrong about that conception.

In recent decades, the connection between consciousness and intentionality is being gradually lost in theoretical writings in linguistics, cognitive science, and philosophy. There has been an effort of varying degrees of explicitness to try to separate the issues concerning intentionality from those concerning consciousness. I think the underlying —and perhaps unconscious— motivation for this urge to separate intentionality from consciousness, even among people who do not share the ideology of the behaviorist-materialist tradition, is that we do not know how to explain consciousness, and we would like to get a theory of intentionality which will not be

---

<sup>4</sup>Freud, S. (1949) pp. 19-25.

discredited by the fact that we do not have a theory of consciousness. The idea is to treat intentionality "objectively", to treat it as if the subjective features of consciousness did not really matter to it. For example, many functionalists will concede that functionalism can't "handle" consciousness (this is called the problem of *qualia*), but they think that this issue doesn't matter to their accounts of belief, desire, etc., since these intentional states have no *quale*, no special conscious qualities. They can be treated as if they were completely independent of consciousness. Similarly, both the idea of some linguists that there are rules of syntax that are psychologically real but totally inaccessible to consciousness and the idea of some psychologists that there are complex inferences in perception that are genuine psychological inferential processes but inaccessible to consciousness, try to separate intentionality from consciousness. The idea in both cases is not that there are mental phenomena which just happen to be unconscious, but somehow in some way they are in principle inaccessible to consciousness. They are not the sort of thing that could be or could ever have been conscious.

I think these recent developments are mistaken. There are deep reasons having to do with the nature of mental phenomena whereby our notion of an unconscious mental state is parasitic on our notion of a conscious state. Of course, at any given moment, a person may be unconscious; he or she may be asleep, in a coma, etc.; and of course, many mental states are never brought to consciousness. And no doubt there are many which could not be brought to consciousness for one reason or another —they may be too painful and hence too deeply repressed for us to think of them, for example. Nonetheless, not every state of an agent is a mental state and not even every state of the brain which functions essentially in the *production* of mental phenomena is itself a mental phenomenon. So what makes something mental when it is not conscious? For a state to be a mental state, and *a fortiori* for it to be an intentional mental state, certain conditions must be met: What are they?

To explore these questions let us first consider cases which are clearly mental, though unconscious, and contrast them with cases which are "unconscious" because not mental at

all. Think of the difference, for example, between my belief (when I am not thinking about it) that the Eiffel Tower is in Paris, and the myelination of the axons in my central nervous system. There is a sense in which both are unconscious. But there is a big difference in that the structural states of my axons couldn't be themselves conscious states, because there isn't anything mental about them at all. I assume for the sake of this argument that myelination functions essentially in the production of my mental states, but even if myelinated axons were themselves objects of experiences, even if I could feel inwardly the state of the myelin sheathes, still the actual structures are not themselves mental states. Not every unconscious state in my brain which (like myelination) functions essentially in my mental life is itself a mental state. But the belief that the Eiffel Tower is in Paris is a genuine mental state. Even though it happens to be a mental state that most of the time is not present to consciousness. So here are two states in me, my belief and my axon myelination; both have something to do with my brain; and both are unconscious. But only one is mental, and we need to get clear about what makes it mental and the connection between that feature — whatever it is — and consciousness.

There are at least two constraints on our conception of intentionality which any theory of the unconscious must be able to account for: It must be able to account for the distinction between phenomena which are genuinely intentional and those which in some respects behave as if they were but which are not in fact. This is the distinction I have discussed elsewhere between *intrinsic* and *as-if* forms of intentionality.<sup>5</sup> And second it must be able to account for the fact that intentional states represent their conditions of satisfaction only under certain aspects and those aspects must matter to the agent. My unconscious belief that the Eiffel Tower is in Paris satisfies both of these conditions. My having that belief is a matter of intrinsic intentionality, and not a matter of what anybody else chooses to say about me or how I behave or what sort of stance someone might adopt toward me. And

---

<sup>5</sup>Searle, John R. (1980), Searle, John R. (1984a) and especially Searle, John R. (1984b).

the belief that the Eiffel Tower is in Paris represents its conditions of satisfaction under certain aspects and not others. It is, for example, distinct from the belief that 'the tallest iron structure built in France before 1900 is located in the French capital', even assuming that the Eiffel Tower is identical with the tallest iron structure built in France before 1900, and Paris is identical with the French capital. We might say that every intentional state has a certain *aspectual shape*; and this aspectual shape is part of its identity, part of what makes it the state that it is.

## 1

These two features, the fact that an unconscious intentional state must nonetheless be intrinsically mental and the fact that it must have a certain aspectual shape, have important consequences for our conception of the unconscious. Specifically, we understand the notion of an unconscious mental state only as a possible content of consciousness, only as the sort of thing which, though not conscious, and perhaps impossible to bring to consciousness for various reasons, is nonetheless the *sort of thing* that could be or could have been conscious. Someone might have mental states which are impossible to bring to consciousness —because of repression or brain lesions or what have you— but if they are genuine unconscious *mental* states, they can't be the sort of thing which in the nature of the case no one could ever have brought to consciousness. To be mental at all, they must be at least possible candidates for consciousness.

The argument for this thesis is a bit complex, but the central idea behind it can be given a simple formulation: the concept of an intrinsic intentional mental state is the concept of something that has an aspectual shape. All representation is under aspects. You can see this, if it is not obvious on its face, by reminding yourself that mental contents are possible or actual contents of *thoughts* or *experiences*. What you can believe, you have to be able to think; and what you can perceive, you have to be able to experience perceptually. But the notions of thinking and experiencing are notions which

imply the presence of aspectual shapes and that in turn implies accessibility to consciousness. The link, then, between intentionality and consciousness lies in the notion of an aspectual shape. To be intentional, a state or process must be thinkable or experienceable; and to be thinkable or experienceable, it must have an aspectual shape under which it is at least in principle, consciously thinkable or experienceable. It must be the sort of thing that could be the content of a conscious thought or experience.

I will now spell out this argument in more detail. For the sake of clarity I will number the major steps in setting out the argument, though I do not mean to imply that the argument is a simple deduction from axioms. Sometimes, indeed, I will be approaching the same goal from more than one path.

1. The first step is simply to remind ourselves that there is a distinction between intrinsic intentionality and *as-if* intentionality; only intrinsic intentionality is genuinely mental.

I have argued at some length for this rather obvious distinction in other writings (Searle 1980, Searle 1984a, and Searle, 1984b), and I will not repeat the arguments here. I believe the distinction is obviously correct, but the price of giving it up would be that everything then becomes mental, because relative to some purpose or other anything can be treated *as-if* it were mental. E.g., water flowing downhill behaves *as-if* it had intentionality. It *tries* to get to the bottom of the hill by ingeniously *seeking* the line of the least resistance, it does *information processing* in order to *calculate* the size of rocks, the angle of the slope, the pull of gravity, etc. But if water is mental then everything is mental.

2. The second step is to point out that the distinction between the *intrinsic* and the *as-if* lies neither in the behavior of the system in question nor in the fact that there are underlying causal mechanisms which cause the behavior nor in the complexity of the system. These features are not sufficient to account for the distinction, since they can be duplicated equally in both intrinsic and *as-if* systems.

3. The distinction does lie, at least in part, in the fact that the idea of an intentional state is closely connected to the ideas of thinking and experiencing. To be a possible belief something has to be a possible thought content: It is a



conceptual truth that anything that can be believed can be thought. Similarly anything that can be a perception can be the content of a perceptual experience. Of course there are plenty of *unconscious* thoughts but even if unconscious they still have to be thoughts.

4. Thoughts and experiences and hence intrinsic intentional states generally, have a certain sort of aspectual shape. They represent their conditions of satisfaction under aspects. For example, the desire for water can be a different desire from the desire for  $H_2O$ , even though there is no way to satisfy the one without satisfying the other. (This aspectual shape of intentional states explains the referential opacity of ascriptions of intentional states. The ascriptions are opaque because the states themselves are aspectual.<sup>6</sup>)

5. This aspectual feature must matter to the agent. It must exist from his/her point of view.

It is, for example, from my point of view that there can be a difference for me between my wanting water and my wanting  $H_2O$ , even though the external behavior that corresponds to these desires may be identical in each case. In the case of conscious thoughts, the way that the aspectual shape matters is that it is constitutive of the way the agent thinks about a subject matter: I can think about my thirst for a drink of water without thinking at all about its chemical composition. I can think of it *as* water without thinking of it *as*  $H_2O$ .

Its obvious how it works for conscious thoughts and experiences, but how does it work for unconscious mental states? One way to get at the main question of this discussion is to ask, "What fact about an unconscious intentional state gives it the particular aspectual shape that it has, i.e. what fact about it makes it the mental state that it is?"

6. The aspectual feature cannot be exhaustively or completely characterized solely in terms of third person, behavioral, or even neurophysiological predicates. None of these is sufficient to give an exhaustive account of the way it seems to the agent.

---

<sup>6</sup> See Searle, John R. (1983). Chapter 7, for an extended discussion of this issue.

Behavioral evidence concerning the existence of mental states, including even evidence concerning the causation of a person's behavior, no matter how complete, always leaves the aspectual character of intentional states underdetermined. There will always be an inferential gulf between the evidence for the presence of the aspect and the aspect itself. (This is one feature of the other minds problem.)

It is less obvious that a similar condition holds for neurophysiological facts, but it does. Since the neurophysiological facts are always causally sufficient for any set of mental facts<sup>7</sup> someone with perfect causal knowledge might be able to make the inference from the neurophysiological to the intentional at least in those few cases where there is a law-like connection between the facts specified in neural terms and the facts specified in intentional terms. But even in these cases, if there are any, there is still an *inference*. The specification of the neurophysiological in neurophysiological terms is not yet a specification of the intentional, even though the neurophysiological is causally sufficient to fix the mental.

Any account that uses only a third person objective vocabulary will leave their aspectual character underdetermined; because no third person objective vocabulary, by itself, will be sufficient to characterize all of the aspectual facts. Thus to take an example of behavioral evidence, the fact that a person consumes H<sub>2</sub>O only if it is described to him as "water" and not if it is described as "H<sub>2</sub>O" would reasonably lead us to conclude that the person desires the substance under the aspect, "water" and does not desire it under the aspect "H<sub>2</sub>O". But this is still an inductive inference; the behavior still underdetermines the aspect in that it is still in principle possible for the person to exhibit exactly that behavior and yet represent his own desires to himself in ways that are not completely manifested to us in his behavior.

---

<sup>7</sup> For these purposes I am contrasting "neurophysiological" and "mental", but of course on my view of mind body relations, the mental simply is neurophysiological at a higher level ( see Searle, (1984a)). I contrast mental and neurophysiological as one might contrast humans and animals without thereby implying that the first class is not included in the second. There is no dualism implicit in my use of this contrast.

To take the more extreme case of neurophysiological evidence, imagine that we had a perfect science of the brain that enabled us to conclude with empirical certainty from the fact that a man was in state S that he wanted a substance under the aspect "water" and not under the aspect " $H_2O$ ". All the same the characterization "S" is still not yet a characterization of those aspects. S will both cause and realize the aspects but under that description it is still not constitutive of the aspects.

This point is so important that it is worth digressing briefly to explain some of its implications. No amount of purely third person behavioral evidence is sufficient to entail all of the aspectual features of intentional states. Such third person evidence always leaves the aspectual features underdetermined. This is one of the weaknesses of behaviorism as a thesis in the philosophy of mind. Behaviorism in the philosophy of mind is now generally regarded as discredited, however, it survives vestigially in the philosophy of language. And you see it in an extreme form in Quine's writings. If you think, as he does, that all of the facts that there are about linguistic meaning are public third person facts, and if you think, as again he does (and as again I believe is mistaken), that the principle that language is public implies that all the facts that there are about meanings are third person facts, then you get not merely the underdetermination of aspectual facts by the behavioral facts but you get an *indeterminacy* of aspectual facts. Thus, Quine argues that within certain limits there simply is no fact of the matter about under what aspects a speaker represents a situation when he utters an expression that we are inclined to translate as "rabbit". I have argued elsewhere that this view is simply a *reductio ad absurdum* of linguistic behaviorism. Since it is obvious from my own case that there are determinate aspectual facts, if behaviorism leads to a denial of this point that can only show that behaviorism is false.<sup>8</sup>

This is not the place to go into detail, but I believe similar remarks can be made about functionalist analyses of inten-

---

<sup>8</sup>Searle, John R. (1987).

tional states, at least insofar as they try to give a third person, objective, account of intentionality. Quine's inadvertent refutation of behaviorism can be applied *mutatis mutandis* to functionalism. Quine's argument shows that any account of purely third person, objective phenomena given in functional terms will always leave the precise details of aspectual facts underdetermined.<sup>9</sup>

Now, these six considerations lead to our first significant conclusion:

7. There is a sense, then, in which intentional states, conscious or unconscious, are irreducibly *subjective*.

The aspectual character is irreducibly subjective in the sense that no characterization in purely neutral third person terms will ever be sufficient to express how the aspectual character seems to the agent, but how it seems to the agent is essential to its identity.

So far so good. We seem to have labored heavily to arrive at the common sense conclusion that there is something subjective about mental states even when they are unconscious. But this leads to a very puzzling question: how could unconscious intentional states be subjective if there is no subjective feel to them, no "qualia", no what-it-feels-like-for-me to be in that state? Indeed how could the ontology of the unconscious be anything other than completely objective? To begin to probe this question, let us ask what is the ontology of ordinary mental states, such as my belief that the Eiffel Tower is in Paris, when I am not thinking about such beliefs. And to avoid muddying the waters, let's leave out halfconscious thoughts, peripheral consciousness, nagging but suppressed conscious states and other shadowy phenomena. What is going on, ontologically speaking, when I have a belief that is totally and completely unconscious but nonetheless there? If there are no conscious neurophysiological processes going on in my brain, then the only other processes would be unconscious neurophysiological processes, that is; *the ontology*

---

<sup>9</sup>For this reason functionalism, like the behaviorism that preceded it, remains programmatic. To my knowledge, no functionalist has so far given anything like a plausible analysis of even one intentional state.

*of the unconscious when unconscious is entirely neurophysiological.* The difference, for example, between my unconscious belief about the Eiffel Tower and the unconscious myelinated condition of my axons is not that the one is a feature of my brain in some mental form and the other is purely neurophysiological; they are both purely neurophysiological. You can see this by imagining that the person is totally unconscious, e.g. in a sound dreamless sleep. Now lots of mental attributions are still true of that person, but the only mental *ontology* is neurophysiological. Indeed this point about unconscious states can be put in the form of a general principle:

8. The ontology of unconscious mental states, at the time they are unconscious, can only consist in the existence of purely neurophysiological phenomena. At the time the states are totally unconscious there simply is nothing else going on except neurophysiological processes.

But now we seem to have a contradiction: the ontology of unconscious intentionality is entirely describable in third person, objective neurophysiological terms, but all the same the states are irreducibly subjective. How can this be? I believe there is only one solution to this puzzle. The apparent contradiction is resolved by pointing out that:

9. The notion of an unconscious intentional state is the notion of a state which is a possible conscious thought or experience. There are plenty of unconscious phenomena, but to the extent that they are genuinely *mental* they must in some sense preserve their aspectual shape even when unconscious, but the only sense that we can give to the notion that they preserve their aspectual shape when unconscious is that they are possible contents of consciousness.

This is the main conclusion of this article. But this answer to our first question immediately gives rise to another question: What is meant by "possible" in the previous sentence? After all, I grant that it might be quite impossible for the state to occur consciously, because of brain lesion, repression, or other causes. So, in what sense exactly must it be a possible content of a thought or experience? This question leads to our next conclusion, which is really a further explanation of 9, and is implied by 8 and 9 together:

10. *The ontology of the unconscious consists in objective features of the brain capable of causing subjective conscious thoughts.* Hence when we describe something as an unconscious intentional state we are characterizing an objective *ontology* in virtue of its *causal* capacity to produce subjectivity. But the existence of these causal features is consistent with the fact that in any given case their causal powers may be blocked by some other interfering causes —such as psychological repression or brain damage.

The possibility of interference by various forms of pathology does not alter the fact that any unconscious intentional state is the sort of thing that is in principle accessible to consciousness. It may be unconscious not only in the sense that it does not *happen* to be conscious then and there, but also in the sense that for one reason or another the agent simply *could not* bring it to consciousness, but it must be the *sort of thing* that can be brought to consciousness because it must be the sort of thing that can be the content of a thought or experience.

To summarize: So far I have tried to establish that intentional states, conscious or unconscious, essentially have an aspectual character and this aspectual character is essentially subjective, in the sense that it cannot be exhaustively accounted for in third person “objective” terms. But since unconscious mental states consist in nothing but objective neurophysiological features of the brain how can they have an intrinsic subjective character? I have suggested that the only explanation for this fact is that unconscious intentional states while not conscious are at least potentially so in the sense that they are the sort of things which could be conscious. When we characterize an unconscious intentional state in terms of its aspectual character, we are characterizing a present brain state in terms of its causal capacity to produce a conscious thought or experience.

Paradoxically, the naïve mentalism of my view of the mind leads to a kind of dispositional analysis of unconscious mental phenomena; only it is not a disposition to behavior, but a ‘disposition’ —if that is really the right word— to conscious thoughts.

So the overall picture that emerges is this. There is nothing going on in my brain but neurophysiological processes. Those

processes are capable of generating conscious states, (which are, of course, higher-level features of the neurophysiological systems and hence neurophysiological themselves). But of the unconscious neurophysiological features, some are mental and some are not. The difference is not in consciousness, for they are both, by hypothesis, unconscious. The difference is that the mental ones are candidates for consciousness. That's all. There isn't any aspectual shape at the level of neurons and synapses.

In my skull, there is just the brain with all its intricacy. All my mental life is lodged in the brain. But what in my brain is my "mental life"? Just two things: conscious states (of course, caused by neurophysiological processes and realized in the structures of the brain) and those neurophysiological states and processes that —given the right attendant circumstances— are capable of generating conscious states. There isn't anything else to the ontology of the unconscious.

## 2

I want to illustrate these points further by imagining a case in which we would have a use for the notion of "unconscious pain". We don't normally think of unconscious pains, and many people, I believe, would accept the Cartesian notion that in order for something to be a genuine pain, it has to be conscious. But I think it is easy to invoke contrary intuitions. Consider the following: it is a very common occurrence for people who suffer from chronic pains, say, chronic back pains, that sometimes the pain makes it difficult for them to go to sleep. And indeed, once they have fallen asleep, there sometimes are occasions during the night when *their condition causes them to wake up*. Now, how exactly shall we describe these cases? Shall we say that during sleep there really was no pain, but that the pain began when they woke up and that they were awakened by neurophysiological processes which normally would cause pain, but didn't cause pains because at the time they were asleep? Or shall we say, on the other hand, that the pain, i.e. the pain itself, continued both before, during and after their sleep, but that they were not

consciously aware of the pain while they were asleep? My intuitions find the second just as natural, indeed probably more natural, than the first. However, the important thing is to see that there is no substantive issue involved. We are simply adopting an alternative vocabulary for describing the same sets of facts. But now consider the second vocabulary: on this vocabulary, we say that the pain was for a while conscious, then it was unconscious, then it was conscious again. Same pain; different states of consciousness of that one and the same pain. We might increase our urge to speak this way if we found that the person, though completely unconscious, made bodily movements during sleep which served to protect the painful portion of his body.

Now what exactly is the ontology of the pain when it is unconscious? Well, the answer seems to me quite obvious. What inclines us to say that the pain continued to exist even though unconscious is that there was an underlying neurophysiological process that was capable of generating a conscious state and capable of generating behavior appropriate to someone who had that conscious state. And in the example as described, that is exactly what happened.

But now if I am right about this then it is hard to see how there could be any factual substance to the old disputes about whether unconscious mental states really exist. If you grant my argument so far, then I am unable to see how it could be other than a purely verbal terminological matter, different only in complexity from the issue about the existence of unconscious pains as I just described it. One side insisted that there really are *unconscious mental* states; the other insisted that if they were really *mental*, why then, they must be *conscious*. But what facts in the world are supposed to correspond to these two different claims?

The evidence that the Freudians adduced involved causal histories, behavior, and conscious admissions by the agent—all of which seemed only interpretable on the assumption of an unconscious mental state, which was just like a conscious state except for being unconscious. Consider a typical sort of case. A man under hypnosis is given a posthypnotic suggestion to the effect that he must crawl around on the floor after coming out of the hypnotic trance. Later, when



conscious, he gives some completely extraneous reason for his behavior. He says, e.g. "I think I may have lost my watch on this floor somewhere", whereupon he proceeds to crawl around on the floor. Now we suppose, with good reason I believe, that he is unconsciously obeying the order, that he unconsciously intends to crawl around on the floor because he was told to by the hypnotist; and that the reason he gives for his behavior is not the real reason at all.

But assuming that he is totally unconscious of his real motives, what is the ontology of the unconscious, right then and there, supposed to be? To repeat our earlier question, what *fact* corresponds to the attribution of the unconscious mental state at the time the agent is acting for a reason of which he is totally unconscious? If the state really is totally unconscious, then the only facts are the existence of neurophysiological states capable of giving rise to conscious thoughts and to the sort of behavior appropriate for someone having those thoughts.

Sometimes there may be several inferential steps between the latent unconscious mental state and the manifest conscious intentionality. Thus, we are told, the adolescent boy who revolts against the authority of the school is unconsciously motivated by hatred of his father. The school symbolizes the father. But again, as in the hypnosis case we have to ask, what is the ontology of the unconscious supposed to be when unconscious? And in this case, as in the hypnosis case, the identification of a specific aspectual shape to the unconscious must imply that there is in the neurophysiology a capacity to produce a conscious thought with that very aspectual shape.

Once you see that the description of a mental state as "unconscious" is the description of a neurophysiological ontology in terms of its causal capacity to produce conscious thoughts and behavior, then it seems there could not be any factual substance to the ontological question: Do unconscious mental states really exist? All that question can mean is: Are there unconscious neurophysiological states of the brain capable of giving rise to conscious thoughts and to the sorts of behavior appropriate for someone having those thoughts? Of course neither side thought of the issue this way, but perhaps

part of the intensity of the dispute derived from the fact that what looked like a straight ontological issue —do unconscious states exist?— was really not an ontological issue at all.

I am not sure I am right about this, but it does seem at least *prima facie* that the old Freudian arguments —involving all that evidence from hypnotism, neuroses, etc.— are not so much conclusive or inconclusive as they are factually empty. The issue is not less important for being conceptual or terminological, but it is important to understand that it is not a factual issue about the existence of mental entities which are neither physiological nor conscious.

### 3

This account of the unconscious has a useful consequence that I want to call attention to immediately. An old puzzle about intentional states has to do with the absence of any clear principle of individuation, and this problem is especially acute for unconscious beliefs. How many unconscious beliefs do I have? We don't know how to get started answering that question. Earlier, I said glibly that I had a belief that was unconscious most of the time to the effect that the Eiffel Tower is in Paris. But do I also believe the following?

Station wagons are inedible.

or

Doctors wear underwear.

If someone asked me whether doctors wear underwear or whether station wagons are inedible, I would have no difficulty in answering; but it seems funny to think of them as unconscious beliefs that I have had all along. Whereas it doesn't seem quite as funny to think of the belief about the Eiffel tower as an unconscious belief. Why the difference?

I can so confidently answer questions about doctors, station wagons and Paris because I have a set of capacities realized in my brain that enable me to generate conscious thoughts and hence generate answers to questions in which I express my conscious thoughts. As long as my thoughts are unconscious they consist only in a neuroanatomy and a neurophysiology

that has capacities for generating conscious thoughts and behavior.

The difference between the case of the Eiffel Tower and the other two cases is that I have already had the conscious thought that the Eiffel Tower is in Paris. I learned it at school, I have seen the Eiffel Tower, I climbed to the top, etc. But the other cases are simply manifestations of the capacity of my brain to generate an indefinitely large number of different conscious thoughts. It is hard but not impossible to count speech acts and even to count one's own conscious thoughts insofar as one can individuate them by content. But when it comes to one's unconscious beliefs, the question of counting is of a different sort altogether: I can't count my unconscious beliefs because there isn't anything there to count except in terms of what I have already consciously thought or in terms of what I could consciously think. But the latter class is indefinitely large and does not reflect a fixed preexisting set of mental representations in the brain.

Our ordinary ways of thinking and talking incline us to think of memory as like a big filing cabinet in which we store a whole lot of information in some language or other —as if written on my brain were a whole lot of sentences in English (or in the "language of thought"). And certain researchers in cognitive science have encouraged this conception by postulating that in the brain there are a whole lot of unconscious "mental representations". But as an ontological claim (and how else are we supposed to take it?) that picture is obviously wrong. What we have instead are a whole lot of capacities in the brain for generating conscious thoughts.

#### 4

Let us return to the question that I asked at the beginning of this article: Can we really think of unconscious states as being like submerged fish or like furniture in the dark attic of the mind? I think these pictures are inadequate in principle because they are based on the idea of a reality which appears and then disappears. But in the case of consciousness, the only reality is the appearance. The submerged belief, unlike the submerged fish, can't keep its conscious shape even when

unconscious; for the only reality of that shape is the shape of conscious thoughts. To repeat, *the ontology of the unconscious is strictly the ontology of a neurophysiology capable of generating the conscious.*

## 5

Now oddly enough, this connection between consciousness and intentionality is lost in discussions of the unconscious mental processes in contemporary linguistics, philosophy and cognitive science. Many of the phenomena which are cited as explanatory psychological features simply could not have any *psychological* reality because they are not the sort of things that could be mental states.

This is a central feature of much contemporary cognitive science, and it is disguised from us by the vocabulary. Some of the key terms are, in effect, a set of puns: "information processing", "intelligent behavior", "rule following", and "cognition" are all used in two quite distinct senses, only one of which is genuinely mental. The reason for this is that the authors in question want a third-person objective science but they also want it to be about a mental reality. It is literally impossible to have both of these features, so they disguise their failure by using a vocabulary that looks mental (what could be more mental than engaging in 'intelligent behavior'?) but which has been stripped of any mental content. And they can get away with this because they can claim that the mental reality they claim to be discussing is all "unconscious"; but that expression now becomes the biggest pun of all because there are two completely different sorts of phenomena called "unconscious", unconscious mental phenomena and unconscious phenomena which have no mental reality at all.

## 6

But someone might object: "Well, why does it matter? Why don't we just scrap the old time folk psychological vocabulary

once and for all and get on with the genuine science of cognition? Why does it matter whether or not we use a mentalistic vocabulary to describe those brain processes which cause genuine mental phenomena or whether we confine ourselves to a purely neutral physiological vocabulary?" The short answer is that it is crucial to understanding the character of the processes involved that we have a clear distinction between those which are mental (hence also physiological) and those which are only physiological. This can perhaps be illustrated by the example of the famous studies of what the frog's eye tells the frog's brain.<sup>10</sup> The retina in the frog's eye filters out most of the stimulus before transmitting the rest of the signal to the frog's brain. The processes in the retina do not literally involve any rule following, nor cognition, nor intelligence. They are simply brute physiological processes. However, their effect is crucial for the frog's conscious intelligent behavior. By making it possible for the frog to have certain visual experiences and not others, they make it possible for the frog to eat and survive.

I give the example of the frog, because here I take it the facts are obvious. But the same thing should be equally obvious about those neurophysiological processes in the human brain which enable us to have genuine intelligence, cognition, and rule following, but which are not themselves cases of intelligence, cognition, or rule following. Unless we are clear about this distinction, we have no hope of understanding how the neurophysiology produces its crucial mental consequences; and indeed, many of the disappointments of cognitive science derive from its failure to pose the question in the appropriate terms. We need, in short, to turn Lashley's claim upside down: Roughly speaking, all genuinely mental activity is either conscious or potentially so. All of the other activities of the brain are simply non-mental, physiological processes, some of which produce conscious and unconscious mental processes.

---

<sup>10</sup> Lettvin, J.Y.; Maturana, H.R.; *et al.* (1959).

## 7 REFERENCES

Churchland, Paul, M. (1984). *Matter and consciousness: A Contemporary Introduction to the philosophy of mind*. Cambridge, Mass.: MIT Press.

Freud, Sigmund (1949). *Outline of Psycho-Analysis*. Translated by James Strachey. The Hogarth Press: London.

Jackendoff, Ray (1987). *Consciousness and the Computational Mind*. Cambridge: MIT Press, A Bradford Book.

Lashley, K. (1956). "Cerebral Organization and Behavior." In H. Solomon, S. Cobb, and W. Penfield (eds.). *The Brain and Human Behaviour*, pp. 1-18, Baltimore: Williams and Wilkins.

Lettvin, J.Y.; Maturana, H.R.; McCulloch, W.S.; and Pitts, W.H. (1959). "What the Frog's Eye Tells the Frog's Brain", *Proceedings of the I.R.E.*, 47: pp. 1940-51.

Lycan, William, G. (1987). *Consciousness*. Cambridge, Mass.: MIT Press. Searle, John R. (1980). "Intrinsic Intentionality". Reply to criticisms of "Minds, Brains and Programs," *The Behavioral and Brains Sciences*. Vol. 3, pp. 450-456.

Searle, J. R. (1983). *Intentionality. An Essay in the Philosophy of Mind*. Cambridge University Press. Cambridge.

Searle, John R. (1984a). *Minds, Brains and Science*. Harvard University Press. Cambridge, Massachusetts.

Searle, John R. (1984b). "Intentionality and its Place in Nature", *Synthese*, Vol. 61, D. Reidel Publishing Company. pp. 3-16.

Searle, J.R. (1987). "Indeterminacy, Empiricism and the First Person". *Journal of Philosophy*. February.