Self-awareness and the mind-brain problem

Gilberto Gomes

Philosophical Psychology, 8 (2), pp. 155-165 (1995)

Current address:
Laboratory of Language and Cognition
UENF, 28013-602, Campos, RJ, Brazil
ggomes@uenf.br

Preprint

Abstract: The prima facie heterogeneity between psychical and physical phenomena seems to be a serious objection to psychoneural identity thesis, according to many authors, from Leibniz to Popper. It is argued that this objection can be superseded by a different conception of consciousness. Consciousness, while being conscious of something, is always unconscious of itself. Consciousness of being conscious is not immediate, it involves another, second-order, conscious state. The appearance of mental states to second-order consciousness does not reveal their true nature. Psychoneural identity can thus be considered a valid hypothesis. Related views of Kant, Freud, Shaffer, Bunge and others are considered. “Naïve psychical realism” is criticized. Consciousness of mental events is considered as the result of the action of a cerebral system that observes the neural events hypothetically identical to mental events. The theory combines a materialist view with a due consideration of subjective experience.

I. The prima facie evidence, afforded by self-awareness, that psychical phenomena are utterly different from physical phenomena seems to be one of the main reasons for the rejection of psychoneural identity thesis. This thesis maintains that psychical phenomena are the working of certain neural systems. The way our mental processes appear to our own consciousness, however, makes this thesis strongly counter-intuitive.

This objection to the identity thesis has been presented by many authors in different ways. Leibniz (1714) stated that if we could go into a brain, as into a mill, we would only see the interaction of material elements, but no perception, thought or emotion. To Hughlings Jackson (1887) it seemed evident that "states of consciousness (or synonymously states of mind) are utterly different from nervous states of the highest centres". According to Köhler (1961), "the events which occur in our organisms appear to us to be of one kind, and those that we call "mental" events to be of another kind. It is this prima facie dualism which makes us speak of a mind-body problem". Another author puts it this way: "however we try to overcome
or evade it, the undeniable facts of conscious awareness and its utter incongruity with the common conception of matter seem inevitably to confront us with an irreducible dualism, either of substances, attributes or relations" (Burt, 1963). Shaffer (1968, p. 49) gives a slightly modified version of Leibniz's argument: "if by X-rays or some other means we were able to see every event which occurred in the brain, we would never get a glimpse of a thought. (...) All we could ever observe in the brain would be the physical events which occur in it." Apter (1970) has also imagined a theoretical experiment in which a sceptic would have his brain exposed while he was awake, in the manner of W. Penfield, and appropriate means would be given him of observing what was happening in his own brain while having a conscious experience. In this way, by confronting the conscious experience with the observed brain events--that is, by extra-linguistic and direct evidence--, he would convince himself of the mind-brain duality. We come to Karl Popper, who states:

we are faced with a prima facie dualism (...) [A]re we able (...) to provide an adequate alternative explanation of our prima facie dualism? (...) I will suggest that the theories produced by materialists to date are unsatisfactory, and that there is no reason to reject our prima facie view (Popper & Eccles, 1977, p. 51).

I will try to show that this objection to the identity thesis may be superseded by a different conception of consciousness. Rosenthal (1986) has also criticised the usual Cartesian concept of consciousness and supported a different one, which is essentially the same I have favoured (Gomes, 1982) as offering a solution to the objection presented above.

II. What we are confronted with, in the reasoning of all the authors who cling to the prima facie dualism of mental and physical phenomena, is the ability of our own mind to observe itself. This is the crucial point in relation to which I will maintain that a conceptual clarification is needed.

Sartre, in an early phenomenological work (1936), distinguished first-order (irreflexive) consciousness and second-order (reflexive) consciousness. His starting point is the basic contention of phenomenology that consciousness is always of something. Sometimes, consciousness may take some conscious state as its object -- this will be second-order consciousness. My point is that consciousness, while being conscious of something, is always unconscious of itself. A conscious state is never conscious of itself, it is always conscious of something else. It may be taken itself as an object of consciousness, but then we will ipso facto have a second process of consciousness. The latter will be conscious of the former, but in its turn unconscious of itself. This is a departure from our habitual mode of thinking. We tend to consider that, when we are conscious of something, we are immediately conscious of being conscious of it. We usually think that by the same movement of consciousness we are
conscious of something and conscious of being conscious of it. My contention is that, although we may in fact be conscious of being conscious of it, this is a second conscious state, superimposed upon the first. Each conscious state is conscious of something but unconscious of itself.

We should therefore distinguish between consciousness as the process of apprehending something in the conscious mind and consciousness as one of the possible objects of such an apprehension. We usually imagine that we have an immediate access to the former, and it is relying upon this assumption that, after confronting our conscious states with any possible physical (neurophysiological) process, we give the verdict of their essential dissimilitude, of their essential duality. The situation changes if we admit that we have no immediate access to the process of becoming conscious of something, if we reconcile ourselves with the seemingly paradoxical assertion that the state of consciousness is unconscious of itself.

This thesis, however, is not in our habits of thought. Sartre himself, who makes the distinction between first-order and second-order consciousness, considers that first-order consciousness is conscious of itself at the same time it is conscious of something. According to him:

consciousness is conscious of itself (...) in so far as it is conscious of a transcendent object (...) consciousness is purely and simply consciousness of being conscious of this object (...) It must be added that this consciousness of consciousness -- not included in the cases of reflexive consciousness on which we will presently insist -- is not positional, which means that consciousness is not to itself its object (1936, pp. 22-23).

All consciousness is at the same time positional consciousness of an object and non-positional consciousness of itself (1943, p. 19).

Well, I would say that consciousness is nothing more than ‘consciousness of its object’, not that it is "consciousness of being conscious of this object". I see no sense in saying ‘consciousness of consciousness’ and at the same time not considering the second term as referring to the object of the first. If I say ‘consciousness of ...’, it seems to me that any term that I use to complete the expression will refer to the object of consciousness. I cannot in fact see what meaning the expression ‘non-positional consciousness of itself’ can have, except as a means of escaping the apparent contradiction of considering consciousness as unconscious of itself. This seems to be an unacceptable paradox within the phenomenological position, which limits itself to what is given by consciousness. But it is no longer so if we go beyond the limits of this position, and consider consciousness, according to the identity hypothesis, as the working of some neurophysiological systems.

In the history of philosophy and psychology, the distinction between the act (or state) and the content of consciousness, between ‘sensing’ and ‘sensum’, between the event of
perception and the percept, and so on has often been emphasized. And, as we will soon discuss, Kant has very keenly stressed that mental states cannot be known as they are in themselves, but only as they appear to inner sense. But these insights have not precluded the prevalent assumption that the state of conscious experience is transparent to itself, that awareness of something is at the same time awareness of this awareness itself.

To my mind, then, as there is no such thing as 'non-positional consciousness of itself', Sartre's concept of first-order consciousness is in fact an amalgamation of first-order and second-order consciousness. The term 'reflexive consciousness’ is also somewhat misleading, as it obscures that two states of consciousness are involved. Besides, when we speak of self-awareness, we should have in mind that this ‘self-' refers to the person, not to the state of awareness itself. In other words, what we should mean is that the person is aware of something happening in his own mind.

III. It is perhaps due to the rapidity of the succession of "acts" (or states) of consciousness (making up what William James called the "stream" of consciousness), and perhaps also to their simultaneous occurrence, that we tend to amalgamate consciousness of something with consciousness of being conscious of it, that we tend to consider them as one and the same thing. It is as if consciousness were transparent to itself, as if knowledge about our consciousness were something given to us without any mediation. Thinking this way, we compare consciousness of our own mental states with what we know or can imagine about the functioning of our brain, and we see no possible identity. Indeed, if the state itself of having awareness of something were transparent to our knowledge, if the perception we have of our conscious processes represented a direct and immediate knowledge of them, we could not fail to acknowledge their singularity as one kind of component of the world. It would then be difficult--unless we chose not to pay attention to our own consciousness--to escape some conception of a res cogitans or a "second world" (as K. Popper proposes (Popper & Eccles, 1977)) radically different from the world of material reality.

If, on the other hand, we admit that the state of being conscious of something is not apprehended by this state itself, the idea that it is the occurrence of some neural processes will become plausible. If we have no immediate access to the process of bringing something into consciousness, we will no longer try to assess the proposed identity directly from the data of consciousness. This identity will appear to us as a possible hypothesis, not as an evident impossibility. The identity to neural states will be a hypothesis concerning something about which we have only mediated knowledge, since in order to be conscious of it we need a second process of consciousness. This second-order conscious state gives us only the appearance of the first conscious state, not its true essence.
IV. One could say that this does not solve the problem at all. Even though the process of consciousness may not be immediately accessible to us, it may still be the object of our own consciousness. As stated above, there are second-order conscious states, whose contents are first-order conscious states. This is still a privileged (though not immediate) access to one's own conscious states. And our first-order conscious states, as they appear to our second-order consciousness, still do not look like neural events. But this argument cannot in fact be justified. The states of my consciousness that I am conscious of are objects of second-order consciousness. Our access to them is mediated by the process of bringing them into second-order consciousness. There is a difference between first-order states of consciousness themselves and the way they appear to second-order consciousness. The contents of my second-order states of consciousness are like "percepts" originated from data coming from first-order conscious states. We should not expect identity between these "percepts" and the neural events that originate them, just as we do not presuppose identity between the percept of a table I see, for instance, and the real physical object. When we see a Necker cube "jump" to the other position, it would be naive to suppose there has been a corresponding change on the sheet of paper on which it is printed. Of course it is our percept that has changed, not the real object. The difference between an external thing and its percept is not always evident to a child or to an uncultivated mind. I am holding that the same difference applies to the internal objects of our second-order consciousness.

V. The opponent of identity theory might say that the problem of prima facie dissimilarity has merely been relocated. Instead of mental states themselves, we are now dealing with the appearance of mental states. But if identity theory were true, he would say, these appearance states would have to share all their properties with some neural state, and this does not seem to be the case. The problem would have merely been shifted from mental states themselves to these appearance states. I will show why this objection is flawed.

According to what I am proposing, there are two states of consciousness involved in consciousness of being conscious (so-called reflexive consciousness). According to the identity hypothesis, this must correspond to two different sets of neural events. So we must conceive that the perception of the table is one set of neural events (\(A\)), that uses information received (through the eyes) from the light reflected by the physical object that is the table. And being conscious of perceiving the table will be another set of neural events (\(B\)), that uses information received from the first set (\(A\)).

Becoming conscious of perceiving the table must indeed be a neural process (process B), but since we are not at this same stage conscious of this process, we have no way of assessing the plausibility of this proposed identity through self-awareness. When I become
conscious of perceiving the table, what I have in consciousness is the appearance of my perceiving the table (the percept I have of process A), not the process itself of becoming conscious of this appearance. What my version of identity theory implies is not that the appearance of a mental state (percept of set A) should share all its properties with some neural state, but that the state of being conscious of this appearance should (hypothetically) be a neural state (set B) --just as the mental state that gives rise to the appearance is (hypothetically) another neural state (set A).

(The appearance of mental states should not be confounded, as in the objection above, with "appearance states", that is, states of being conscious of the appearance of a mental state. The state of being conscious must not be confounded with the content of consciousness.)

VI. Kant (1787) has stated that mental states cannot be known as they are in themselves, but only as they appear to inner sense (B 55). According to him, we perceive ourselves internally only as we are affected by ourselves (B 69, 152-3). In relation to internal intuition, we only know our own subject as a phenomenon, not as it is in itself (B 68). All experience, the internal experience not less than the external, is knowledge of objects only as they appear to us, for it depends not only on the character of the object but also on that of the subject and its receptivity (1798, § 7, p. 141). Kant also states that, although the I is one in its substance, it is double according to its form: (1) the I as subject of thinking; and (2) the I as object of perception (1798, § 4, footnote, p. 134). According to Kant, I do not know myself as I am, but simply as I appear to myself (B 155-6, 158).

These views of Kant's gain a new relevance when considered in relation to the thesis I am proposing. Of course Kant himself would not have accepted psycho-neural identity. Besides, there is for him no possible human knowledge of things as they are in themselves. Thus, a neural theory of the mind could not be viewed, in Kantian terms, as a theory about what mental events are in themselves. However, his insight that we only have access to our mental processes as they appear to ourselves is highly suggestive from our point of view. It seems to undermine the objection to identity theory based on the prima facie dissimilarity of mental events in relation to physical events, that is, based on self-awareness. If self-awareness only informs us about mental events as they appear to us and not as they are, we can consider a hypothesis about the hidden nature of these mental events, even if this hypothesis cannot possibly refer to what they are, in Kantian terms, in themselves. This hypothesis cannot be ruled out by the mere phenomenal appearance of mental events.

VII. Freud's conception of consciousness and of psychical phenomena goes halfway to the thesis I am proposing. We know that Freud (following Lipps) admitted a radical
dissociation of what is psychical and what is conscious. It is well-known that he admitted the existence of unconscious psychical processes, but it is less well-known that he proposes a dissociation between what is conscious and what is psychical even in relation to conscious phenomena themselves. From a careful reading of the relevant passages, however, it becomes clear that, for him, consciousness is not an intrinsic attribute of the processes of which we are conscious. In a conscious thought, for example, he dissociates two components: the thought proper, which, according to him, is in itself unconscious, and the perception of this thought by consciousness (Freud, 1900, pp. 612-3; 1905, p. 147-8; 1915, p. 171). This is not to say, however, that thought would proceed in just the same way if it were not perceived by consciousness. On the contrary, perception by consciousness provides invaluable feedbacks to the process of thought, which alter its course and represent an essential contribution to it.

The other half of the way to arrive at what I am proposing consists in extending this conception to consciousness itself. Just as Freud said that thoughts and wishes are in themselves unconscious (but can, in some cases, be perceived by consciousness), we could say that the perception itself of a thought or wish, by consciousness, is in itself unconscious. This statement seems paradoxical because of the ambiguity of the words ‘conscious’ and ‘unconscious’, which can have both an active sense (being conscious of something) and a passive one (being consciously known). In the active sense, to say that a state of consciousness is unconscious would obviously be self-contradictory. Of course a state of consciousness is always conscious of something. But in the passive sense, it may be conscious (consciously known) or not. Indeed, perception by consciousness can be said to be in itself unconscious (in the passive sense), since it is not necessarily the object of another process of conscious perception (though it may become so). In other words, the process of becoming conscious of a thought or wish is not in itself consciously known--it may become so, but only through another process of perception by consciousness.

VIII. Some authors who favour the identity thesis tend to escape the problem of prima facie duality, rather then try to solve it or at least face it. D. O. Hebb, for example (who could be said to adopt a central-state theory of psychical phenomena), considering as a logical premiss that self-observation is impossible, arrives at the surprising conclusion that all psychical phenomena are unconscious (1966). (I would agree if he had said: "unconscious of themselves"). He will try to build his psychology in terms of observable behaviours and hypothetical central states or processes, and will leave aside the "internal" observation of psychical states or processes by the subject himself. But to say that all psychical phenomena are unconscious amounts to a sheer denial of what is evident: that we may be conscious not only of external, objective events but also of our own mental states. These (or rather, some of these) are
as observable as any external fact. They are not publicly observable; but to ourselves they are evident, and they may be communicated to others.

Hebb's position implies a rejection of what Shaffer calls the "first-person account of consciousness". According to Shaffer (1968, p. 21), the "first-person" point of view has to be taken into account in relation to consciousness. Conscious states are not just internal states that we attribute to "him", to "her" or to "them" (this would be the "third-person account of consciousness"). We must also consider the "first-person account", that is, the account each one of us may give of his or her own conscious states. And I believe Shaffer is right in stressing the value of this first-person account. Indeed, the data of self-awareness are precious to the understanding and investigation of psychological phenomena, and to the building of psychological theories.

Bunge (1979), among others, evades the problem that this first-person account poses to the identity thesis (that is, the problem of prima facie duality) by considering it a linguistic problem. According to him, ordinary language favours dualism and, if we turn to the language of science, the problem disappears. "In science (...) we do not reify properties, states or events. (...) Wind is the motion of the air: there is no wind apart from moving air. [...] Similarly, there are (...) only minding bodies. (...) the mind is (...) a collection of functions (activities, processes) of certain neural systems (...)" (pp. 124 and 183-4). In science, we do not reify the motion of a physical object, for instance, a ball. The thing is the ball, the motion is what happens to it. Likewise, according to Bunge, we should not reify conscious phenomena. This goes very well for the third-person account of consciousness. Anxiety in another person, for instance, may be viewed as what happens to certain neural systems of his, causing some forms of behaviour. The problem is that in myself--and this applies to other people's selves too, according to what they tell me--I can perceive the anxiety I feel as much as I can see the ball. And, what is more, in the perception of this anxiety I do not perceive any neural systems as such, nor anything that seems analogous to what I can conceive as the activity of neural systems. So the problem is real, not a false one due to our linguistic habits. Hunger, love or the understanding of the operation of multiplication are not just states or processes that we may attribute to other people's or our own neural systems, considering them as influencing their or our behaviour. They are also something we are aware of, in ourselves. Bunge himself recognizes that "when one thinks about something, he does not feel his thoughts to be located anywhere" (pp. 144-5). And we could add: he does not feel that his thoughts are a state or process that happens to his brain.

Second-order or so-called "reflexive" consciousness is an undeniable fact and one that poses a real problem to the identity thesis. We may propose explanations of second-order consciousness, from the third-person point of view, in terms of feedback loops or self-monitoring cognitive systems. But, from the first-person point of view, it will still present us
with the *prima facie* heterogeneity between mental and physical phenomena. It is only by ignoring it that we can escape this problem. But it seems to me that we can face it and that the conception of consciousness as unconscious of itself is a solution to it.

**IX.** Observation of reality is of course indispensable to science. But science does not attribute all the phenomenal characteristics of our perception of things to its conception of the physical reality of these things. On the contrary, the rejection of naive realism (that is, considering things as we perceive them as reality itself) is essential to science. The scientist must always admit that there is a "hidden reality" different from or even opposed to the phenomenal appearance of things. The moon seems bigger than the stars but is in fact smaller. Physics tells us that different colours of light correspond to different ranges of the wave length of electromagnetic radiation. But nobody sees the length of the waves of electromagnetic radiation as such. We only see light of different colours. The point is rather obvious in relation to external reality. What is not obvious is that the same reasoning applies to inner experience. Here, too, the phenomenal appearance of a mental event should not be taken to be the mental event itself. The realization that even the *subjective* side of a psychical phenomenon can only be the *object* of conscious experience should prevent us from falling into a "naive psychical realism". We must accept that there is also an "inner hidden (psychical) reality" to be discovered behind the form in which mental states come to our awareness.

It is true that the statement "psychical phenomena are the working of certain neural systems" sounds strange, and is hard to accept at first sight. So is the statement that a solid iron ball is in fact a collection of tiny elementary particles separated by void space, each one at a long distance from the others. The statement that the Earth is round and suspended in space without having anything to prevent it from falling "down" must also have sounded very strange at first; and it still sounds so to any child when he or she is first told about this.

Mental naive realism, then, is the view that the phenomenal appearance of inner mental states is the reality itself of these states. Kripke states that "in the case of mental phenomena there is no ‘appearance’ beyond the mental phenomenon itself" (1980, p. 154). I would say there is; and in so doing, I am making a distinction similar to Kant's distinction of the subject as it appears to inner sense and the subject as it is in itself (Kant, 1787, B 55, 68).

**X.** We could say that consciousness as an object of consciousness is a percept derived from information received from certain *central* neural processes, just as an external object, as I see it, is a percept derived from information received from certain *peripheral* neural processes.

We can now go back to the thought experiments of Leibniz, of Apter and of Shaffer (see section I). If we could go into a brain, we would not perceive any thought or inner
experience, because the way of perceiving them—*as thoughts or inner experiences*—is receiving
the relevant information through a neural channel within one's own brain. Suppose that by
suitable techniques we could observe all neural events happening in a brain (or in our own
brain). The first problem (and not a small one) would be selecting the relevant information,
since our techniques would certainly reveal a wealth of physiological details not relevant to the
psychical processes in question. But even if the neural processes identical to some psychical
phenomenon could be adequately selected, we would still be perceiving them (on the screens of
our equipment, or through loudspeakers) with our vision, with our hearing, and not with our
"endopsychical perceptual system" (second-order consciousness). Our observation of mental
processes of our own can be considered as a perceptual modality of its own kind. One would
not expect to hear music with one's sight, or to see things with one's ears. To "go into a brain",
or to observe it with some apparatus, and expect to perceive thoughts or emotions as such
would be like expecting to hear sounds by looking at a visual display of sound waves.

According to Shaffer, "the appearance to the naked eye of a neurological event is
utterly different from the experience of a thought or pain". In more precise terms, we could say
that the appearance of a neurological event to the eye (naked or not) is utterly different from its
appearance to second-order consciousness. Shaffer states: "It is sometimes suggested that the
physical aspect results from looking at a particular event ‘from the outside’, whereas the mental
aspect results from looking at the same event ‘from the inside’" (p. 45). He objects to this,
saying that the inner aspect of one's brain is probably far more accessible to a brain surgeon
than to oneself. No doubt about that, but what is meant by the rather inaccurate expression
‘looking from the inside’ is that there is a specific "built-in" way of observing the event in
question, and that any other way of observing it will be through one of the sensory systems
whose task is the analysis of information coming from outside the nervous system.

To say that consciousness of our own mental processes can be considered as a kind of
perceptual modality does not imply that we are mere observers of our own thoughts. The mental
subject has surely an active role in thought and conscious experience. The active subject should
not be identified with self-awareness. Observation of one's own mental states is one function of
the mental subject, but not the only one. On the other hand, as we have already seen, the mental
subject itself may become the object of self-awareness. (Or more precisely, some aspect of the
mental subject may become the object of another state of awareness).

Self-awareness is often considered as an attribute, as a *quality* of some processes. I
believe we should consider it as a *process* of its own; it is an active process, not a mere state or
quality.

I believe the theory presented makes it possible to keep the identity thesis without
eliminating or failing to consider the data of self-consciousness. This has the advantage of
combining a materialist view (that interrelates with biological sciences) with a theoretical perspective that permits taking into account the positive contributions of the schools of psychology that make use of the data of subjective experience.

Notes

1. Type-type identity theory considers that mental event-types are strictly identical with physical event-types, and so psychological explanations can in principle be reduced to explanations using physical or biological concepts. A better alternative is Davidson's anomalous monism (Davidson, 1970), according to which mental event-tokens are identical with physical event-tokens, but mental predicates cannot be reduced to physical predicates, since mental event-types do not necessarily correspond to neural event-types. According to this view, neurophysiology may provide important bases for psychological theories, but it will always be insufficient to account for psychical facts, which necessarily require psychological concepts or explanations.

2. The object of this paper is thus sharply circumscribed. Important questions related to the mind-brain problem such as other objections to identity theory, the issues raised by functionalism, etc. are clearly beyond the scope of this paper.

3. In English, the word ‘awareness’ would perhaps be better suited to this meaning.

4. One could raise the question of which mental processes can occur without consciousness. Freud himself admits the existence of thoughts which are not perceived by consciousness. (See, for instance, 1900, p. 593; 1905, p. 160). But the question is not in fact essential to our argument. What concerns us here are conscious thoughts, wishes or any other conscious mental processes. The traditional view considers consciousness as a property of these processes themselves, the property of our having immediate knowledge of them. The alternative view is that they are composed of two interrelated processes: the thought (or wish, etc.) itself, and its perception by consciousness.

5. See James Strachey's comments on this point, in Freud, 1915, p. 165, note 1. According to him, this ambiguity is scarcely present in the German, where ‘bewusst’ and ‘unbewusst’ usually have the passive sense of ‘consciously known’ and ‘not consciously known’.
References


Leibniz, G. (1714) *Monadology*, section 17.


