PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Investigating the biology of consciousness

A. R. Damasio

Phil. Trans. R. Soc. Lond. B 1998 **353**, 1879-1882
doi: 10.1098/rstb.1998.0339

THE ROYAL SOCIETY

Investigating the biology of consciousness

Antonio R. Damasio

Department of Neurology, Division of Behavioral Neurology and Cognitive Neuroscience, University of Iowa College of Medicine, 200 Hawkins Drive, Iowa City, IA 52242, USA

The fact that consciousness is a private, first-person phenomenon makes it more difficult to study than other cognitive phenomena that, although being equally private, also have characteristic behavioural signatures. Nonetheless, by combining cognitive and neurobiological methods, it is possible to approach consciousness, to describe its cognitive nature, its behavioural correlates, its possible evolutionary origin and functional role; last but not least, it is possible to investigate its neuroanatomical and neurophysiological underpinnings. In this brief essay I distinguish between two kinds of consciousness: core consciousness and extended consciousness. Core consciousness corresponds to the transient process that is incessantly generated relative to any object with which an organism interacts, and during which a transient core self and transient sense of knowing are automatically generated. Core consciousness requires neither language nor working memory, and needs only a brief short-term memory. Extended consciousness is a more complex process. It depends on the gradual build-up of an autobiographical self, a set of conceptual memories pertaining to both past and anticipated experiences of an individual, and it requires conventional memory. Extended consciousness is enhanced by language.

Keywords: core consciousness; extended consciousness; cognitive phenomena

1. INTRODUCTION

With a few exceptions to the contrary, consciousness is presumed to be the most complex and impenetrable human property, from which follows that it is the most difficult to define and the most problematic to investigate. For some of those who are preoccupied with investigating the relation between mind and brain—neuroscientists, cognitive scientists and philosophers of mind—consciousness and mind are one and the same, and, as a result, the intangibility, unapproachability and refractoriness of consciousness are those usually associated with the mind. Some of those investigators ask the following question: how can science approach interior phenomena that can be made available only to a single observer and are thus hopelessly subjective? They answer the question negatively and so it is not surprising to discover that the current discussion on the scientific account of consciousness often aligns itself with one of the following positions:

(i) declare consciousness the supreme scientific mystery and propose that none of it can ever be explained;
(ii) declare consciousness a respectable mystery that will be solvable only when a nearly equivalent mystery, for instance quantum gravity, will yield its own solution and permit us to deal with the issue;
(iii) declare a possible scientific solution in the study of externally observable manifestations of this internal phenomenon, in the hope that the allegedly objective study of a part of the thing will be enough to understand the whole.

For a variety of reasons, none of these defeatist positions is really well founded. On the one hand, a somewhat more encouraging position is no more acceptable either. It consists of declaring scientific victory by explaining that consciousness was a mere illusion and that the fundamental explanations are already available.

Although this is not the place to deal in detail with a counter-argument to all of these positions, I shall explain briefly some of my fundamental disagreements.

2. CONSCIOUSNESS IS NOT AN ILLUSION

I shall begin by emphasizing that there is no advantage in considering consciousness as an illusion. The necessarily subjective feeling that we call consciousness is very real, is shared by all of us, writing or reading these words, and it is that reality that requires explanation. It is easy to understand the inclination of a scientist to deal with phenomena that can be observed by others; that is, phenomena that can be reported from a second-person or third-person perspective, rather than dealing exclusively with phenomena observable only from a first-person perspective. In fact, however, when we deal with mind in general and consciousness in particular, nothing else will do but precisely using the first-person perspective. Studying the behaviours of a conscious subject is very valuable and can be correlated with the study of the same subject's mental processes. But studying the behaviour alone is simply not enough and misses the mark. A purely behaviourist approach is quite adequate for creatures presumed to operate largely on the basis of innate response mechanisms and of conditioning, but it can certainly not do justice to the study of individuals who have not just behaviour but also cognition superimposed on top of behaviour.

It is fair to say that some investigators who concede this point wish to address consciousness by studying the

mechanisms by which mental images are generated, attended, or held on-line. However, the focus on image-making alone also misses the mark. It is conceivable, and in fact it is in all likelihood true, that behaving organisms can generate images of varied sensory modalities for the purpose of optimizing motor responses. This does not mean at all that they use images to create a sense of self or to create subjectivity. In other words, there can be a substance of mind without the trait of consciousness. Again, what we must explain if we are to address the issue of consciousness is the generation of a sense of self and the generation of the sense that such self is involved in the process of perceiving the stimulus. Consciousness occurs when we can generate, automatically, the sense that a given stimulus is being perceived in a personal perspective; the sense that the stimulus is 'owned' by the organism involved in the perceiving; and, last but not least, the sense that the organism can act on the stimulus (or fail to do so), that is, the sense of 'agency'.

3. CONSCIOUSNESS IS NOT BASED ON LANGUAGE

An easy way out on discussions of consciousness is the notion that along with the objective study of behaviours all we need to be concerned with is the study of language, since language might well be the source of consciousness. The idea is that language provides a running commentary on other events of the mind and that human consciousness, in the end, is nothing but that commentary. For a variety of reasons this idea is simply not acceptable. The dependence of consciousness on language would logically rule out the existence of consciousness in any non-human species and in infants. Yet, although one ought to be cautious about commenting on what cannot be observed directly, lucid and comprehensive accounts of the biology of complex non-human creatures suggest that they too are likely to have a basic process of consciousness without which it would be difficult to explain some of the objective behaviours that they exhibit.

The notion of language dependence, however, easily runs into an even greater objection. The essence of language coding is the translation of a set of non-verbal representations, that is a 'concept', into linguistic representations, for instance words, signs and sentences. If language were to be the primary source of consciousness, it would have to be true that terms such as 'I' or 'me' would be free-floating novelties that would be the translation of nothing. Of course, that would be a patent absurdity. Terms such as 'I' and 'me', and phrases such as 'I feel pain', are translations of non-verbal concepts that are themselves representative of non-verbal entities and events. They translate in language terms the non-verbal entities 'organism' or 'individual', and the non-verbal event of 'perceiving in a certain manner'.

4. IT IS POSSIBLE TO STUDY CONSCIOUSNESS

The argument that studying consciousness is not possible is also weak. It is true that the study of interior subjective phenomena poses special problems, but the problems can be overcome with many of the techniques currently in use in cognitive neuroscience. It is possible to design experiments in which well-controlled external stimuli consistently produce certain internal states, as verified by reports from previously tested groups of individuals. Investigators can certainly not jump into a subject's mind and check the ongoing events, but they can make well-informed predictions about what is likely to be occurring, or not occurring, given the presentation of certain external stimuli. Moreover, it is possible to relate such internal events to a variety of increasingly precise neurophysiological and even neuroanatomical indices, and it is possible to produce measurements on the variation of such indices. The latter are based on the techniques of electrophysiological recording, either from the scalp or from surgically exposed brain structures, the techniques of functional imaging, using either positron emission tomography or magnetic resonance, and techniques such as magnetoencephalography.

5. CONSCIOUSNESS CAN BE DEFINED

The argument that consciousness is impossible to define is equally weak. It is certainly true that the definition is not easy and that unless special care is taken to define, at the outset of any investigation or presentation, what is really being meant by the term, a considerable confusion will ensue. In the end, most of us have a sense of what is meant by consciousness. Most of us, if pressed, produce a definition of consciousness that is patently recorded in the appropriate entry of any quality dictionary: consciousness is that which permits 'awareness of self and surroundings'; or permits 'the awareness of one's own existence, sensations, thoughts, surroundings'. No dictionary fails to present such a definition, usually at the top of the list of definitions. It is not a major problem to make the point that other meanings, such as for instance 'social consciousness', or 'conscience', are semantically related to the primary meaning but are not primary themselves and are not be the main target of biological investigations at this stage.

I have found it helpful to distinguish between 'core consciousness', which can also be designated by the single word 'awareness', and 'extended consciousness', which can be designated by the word 'consciousness' alone. Both are internal phenomena of the mind but core consciousness is more basic than the extended variety. Extended consciousness depends on core consciousness. Both occur automatically, which means that no amount of willpower can either make them happen or prevent them from happening.

Core consciousness, or awareness, allows a living organism to sense that the contents of its thoughts are its own, that they are formulated in the perspective of the organism, and that the organism can act on those thoughts. This ability does not rely on language nor does it require great intelligence or memory. Obviously all humans have core consciousness but I shall venture to say that so do individuals of many non-human species.

Extended consciousness or consciousness proper surveys a larger canvas of thoughts. Those thoughts portray not just the present state of the organism but also its past and its expected future. They depend on the gradual build-up of an 'autobiographical self', a set of memories of the individual's unique past and expected experiences. Conventional memory is required for the

construction of the autobiographical self, and working memory is required for the extensive display of items recalled from the autobiographical self. Finally, language helps the categorizations that enrich the autobiographical self. Nonetheless, it should be clear that extended consciousness proper performs, for a larger compass of contents, precisely the same deed that core consciousness accomplishes for the simple level: it places mental contents in an individual perspective; it confers ownership of those contents; and it gives the owner the sense that it can act on them.

6. CONSCIOUSNESS IS NOT THE PINNACLE OF BIOLOGICAL EVOLUTION

Let me now turn to the argument that regards consciousness as the pinnacle of biological complexity. This argument, incidentally, is interwoven with other arguments such as those that consider consciousness as the ultimate mystery, or that propose that consciousness is impossible to grasp because understanding consciousness would serve no purpose evolutionarily and, accordingly, we would lack the evolutionarily developed mental modules necessary to grasp it. I have a variety of comments on these arguments.

First, I see no reason to give up on neuroscience yet as the source of explanations on consciousness. Not only is it true that we have not exhausted the possibilities of explaining consciousness in neuroscientific terms, but it is also true that we have barely scratched the surface of neuroscience in terms of such an attempt. Second, I find it premature to turn to explanations depending on yet other mysteries, for instance, explanations that depend on quantum physics and especially those that rely on hitherto undeveloped particulars of quantum physics, such as quantum gravity. As I see it, the arguments available in the literature in this regard are well intentioned but somewhat naive. Leaving aside the enormous uncertainty as to whether quantum-level effects trespass across neural levels to have an impact on the behaviour of systems, I have no reservations and concern over how mental phenomena are defined in the related studies. For instance, it is apparent that the physicist Sir Roger Penrose, in his intriguing books and articles, is not really addressing the issue of consciousness *per se* but rather the issue of mind in general. Seen in this light, his claim, and that of others, could be paraphrased as follows: it is not possible to give a full account of mind processes without relying on quantum-level explanations. This is actually quite different from saying that quantum physics is specifically necessary to explain consciousness. In all candour, at this point in the history of studying the neuroscience of mind, it is not reasonable to disqualify the claim that they paraphrased. It might be that a full account of the physicality of mind requires the kind of conception of matter that quantum physics brought to scientific understanding earlier in the century. However, it does not seem reasonable at this point to invoke quantum physics when we are dealing with the specific mechanisms behind the generation of consciousness, that is, behind the generation of the sense of ownership, perspective and agency in mind.

Perhaps most importantly in this brief list of counter-arguments is to see consciousness not as the most complex biological state attainable but rather as a set of adaptive, mid-level biological mechanisms of mid-complexity that are necessary for the survival of complex organisms in a complex environment, and that, in turn, come to permit plenty of other more complex biological states. One might ask, quite reasonably, to which more complex states I refer; I am referring to the creations that arise in the logical manipulations of knowledge that could not occur without a conscious mind. I am referring to the creations that permit the development of rules for social behaviour; the development of a corpus of ethics; the development of laws; the development of the sciences and of technology; and the cultivation of the arts. Consciousness permits all of this, none of those creations being possible in non-conscious individuals. However, all of those creations require a far more extensive and complex biological state than consciousness itself, as witnessed by the multiplicity of systems involved in the requisite manipulations, the compass of obligate knowledge and the temporal dimension of the creative process.

7. CONSCIOUSNESS: HOW AND WHAT FOR?

The central issue for the elucidation of consciousness, then, is this: how do the images that constitute the normal thought process become the property of the organism that is engendering them? To put it in other words, how does the organism generate a sense of observer relative to images that are sensed as observed? Without falling into the well-known trap of accounts of consciousness that invoke the homunculus, infinite regress and a central theatre, I have proposed that the answer to these questions requires the understanding of how the brain can build a representation of 'core self' and how knowing is attributed to such a transiently constructed representation.

I have suggested that the self is grounded on a representation of the organism, that is, on continuously updated representations of the structure and states of the body. Moreover, I have suggested that the process of perceiving depends on transient changes in the representation of the body state that occur as a result of processing images about whatever object we are to place in consciousness.

Perceiving a given object, out in our environment or recalled in our thoughts, engenders a number of responses in our body, not the least of which concerns the emotional state that inevitably accompanies the process and is based on the object's value to the organism—as set innately or as acquired in prior emotionally laden experiences. In this perspective, consciousness becomes inextricable from the mechanisms of life regulation.

Living creatures such as we are, produce core consciousness when our organisms construct images of a part of themselves forming images of something else. Core consciousness occurs when organisms are, so to speak, caught in the act of representing themselves when they represent other things. They, that is we, perform the core consciousness trick because of: (i) its complicated interaction between the body-proper and the brain, (ii) the multilayered structure of the brain, and (iii) the brain's inherent obsession with building images that represent the organism, which include the body-proper

and parts of the brain itself. The images that constitute the representation of part of the organism in action describe how the organism has been 'caught in the act of making other images'. The basic constituent of the images in the set is the internal state of the organism. The caught-in-the-act description is made in a non-verbal language, a language of body states, of somatosensory information.

Consciousness in general, and core consciousness in particular, is the means by which the organism indicates to itself that it is engaged by some object or event. The object or event can occur outside the organism such as an object you can touch or a scene involving varied objects, or it can happen inside the organism—e.g. a pain in the chest. The object or event can be directly perceived or indirectly recalled from the memory of past perceptions. No matter what, inside or outside the organism, directly perceived now or indirectly perceived in recall, the business of core consciousness is to indicate that the continuing processing of images of anything is happening within the individual organism, in its perspective, owned by it. In the most general terms, core consciousness consists of having the organism reveal to its imaged reference, the core self, that it is engaged in the process of making mind. The revelation is transient, pulse-like, and incessantly generated for any object that engages the organism, be it actually present or recalled from memory.

I do not think the ability to make this revelation happened by design, or that it was meant to happen. I believe instead that it was bound to happen as soon as brains had a capacity to represent their own involvement in the process of representing. Core consciousness was born once the presence of certain neuroanatomical devices permitted a description, by one part of the brain, of what other parts of the brain were doing.

Core consciousness might have been born without prior design, as no doubt occurred for so many of the biological devices and processes that we carry today. However, the reason why it prevailed probably came from the advantages that it gave the organisms that had it. And what was that advantage? It was the possibility of connecting the very core of life regulation with the processing of images. Put in other words, it was the possibility of bringing the inherent value system of life regulation to bear on the processing of the images that represent things and events inside and outside the organisms.

Why was that an advantage? It was an advantage because survival in a complex environment, that is, efficient management of life regulation, depends on taking the right action and making the right plan; that, in turn, requires a purposeful manipulation of images in mind.

In this perspective, core consciousness is the door to a revelation of regulatory values, the passage into the possibility of constructing in the mind some counterpart of the regulatory value hidden in the brain core, some new and more open way of sensing the life urge and the means to hold on to life. Core consciousness is the rite of passage that allows an organism armed with reflexes and conditioning to become a minded organism, an organism in which responses are shaped by a naturally born concern with the organism's own life.

When core consciousness began, millions of years and many species ago, we were very far from the current sophistication of modern consciousness, very far from the ease with which we can describe, using language, the reasons behind our actions, past or intended. However, when core consciousness began, we were on the right track and we transcended the critical threshold. We were telling ourselves, without using any words, the answer to the question we never asked, that yes, there was an individual perspective to our percepts, and yes, there was an individual ownership of images, and yes, it was all tied to life.

In its simple, and non-languaged way, core consciousness was and is now saying, in the mind of an organism, that attention must be paid to the images in the rest of the mind of that organism, that attention must be paid because those images probably concern the organism's future.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES