Reprinted from Psyche: problems, perspectives Vol. 1 (2001) 93–100

# Computational aspects of consciousness*

**Jeremy J. Ramsden**

*Faculty of Natural Philosophy, University of Basel, Switzerland*
*and*
*Collegium Basilea, Institute of Advanced Study, Basel, Switzerland*†

**ABSTRACT**

It can be plausibly established, independently of any preconceived dogma, that (i) the brain is an automaton with state structures; (ii) state structures could encompass the total thinking ability of the brain; (iii) mind is an aspect of the thinking power of the brain, therefore mind is an aspect of state structure. There is moreover no particular correspondence between state structure and physical structure. Furthermore no identity, only correlation, has been shown to exist between mental events (thoughts) and material events (neural processes). Thinking is embodied in brain activity but is not same thing. Humans think and understand as agents, and agency may be embodied in biological or artifical structures, yet in neither case can thought and understanding be attributed to the structures. This notion is reinforced by the fairly easy demonstration (via examples from mathematical reasoning) that thought is non-algorithmic. Consciousness does not arise in but is rooted in the animate, and the link between our corporeal conscious experience and actual life appears primordially through proprioception—perhaps the most vital biological activity.

To comprehend the nature of human consciousness is undoubtedly a tremendous challenge; on the one hand one cannot acquire knowledge or understanding without consciousness; on the other, consciousness belongs in the private and unobservable world of subjective mental states (qualia).

## METAPHORS FOR THE BRAIN

The method whereby brain produces mind has for centuries been discussed in terms of the most complex contemporary metaphors of science and engineering. Fifty years ago, this meant the digital computer, just then appearing on the scene, and serving as a powerful stimulus to the development of the idea that the brain is in some ways "merely" a computer, albeit a very powerful and sophisticated one. Such ideas lead to the extreme reductionist view known as "strong artificial intelligence", which asserts that all mental events are reducible to an algorithm, and all brain functions, including consciousness, in principle may, and hence presumably will, occur in computers.

Computation may be defined as the manipulation of symbols, or the processing of information, and undoubtedly both computers and the brain carry out these functions. One concrete result of this link between engineering and biology has been the field of *neural networks*, which deals with the computational properties of networks of "neuronlike" elements, lying somewhere between a model for neurobiology and a metaphor for how the brain computes. Circuits of model neurons have been tested on difficult real-world problems, such as spoken word recognition. If the neural circuit, with its distinctively biological feature, is capable of solving such a problem which circuits without that feature solve poorly, then is it plausible that that feature is computationally useful in biology? This is at best only a weak argument, but one that has been helpful in trying to discern among the mass of details in neurobiology what is truly important, as opposed to merely true. In any case, artificial neural circuits have already been put to practical use for solving difficult problems, such as in the chemical analysis of complex mixtures (e.g. [1]).

Many of these kinds of problem fall into the class of pattern recognition—something at which standard digital computers are remarkably bad, and we, along with all living creatures, are remarkably good—not surprisingly, for it is of essential importance for survival. The inability to distinguish edible mushrooms from poisonous toadstools could have fatal consequences, and the inability to recognize members of the opposite sex might lead to the failure of the procreative system so necessary for ensuring the survival of our race.

At the same time as work has advanced with artificial neural networks, simulated on silicon-based computers, the perception has grown that the brain is not at all like any computer constructed by man. The physical divergence is of course obvious even from a perfunctory investigation of the insides of a computer and a human brain. Apart from the fact that the basic elements of a modern digital computer (transistors etc.) behave very differently from the basic elements of a brain, i.e. the neurons, the ways they are interconnected are also very different. Impressive as modern very large scale integrated circuits are, they come nowhere near the almost unimaginable complexity of a human brain: each of its ten or so milliard neurons (each of which is itself a little computer, albeit a relatively simple one, seemingly operating on the 'integrate and fire' principle) is connected to several thousand others, making a total of the order of $10^{13}$ –$10^{14}$ —several *tens of billions*—of connexions.

*Based on a lecture to all Faculties delivered on 20 December 2000 at the University of Basel.

†Correspondence to: Hochstrasse 51, 4053 Basel, Switzerland. E-mail: J.Ramsden@unibas.ch

In the study of the (human) brain, it has rightly been emphasized that one must take its biological features into account, including the way the brain is formed, both in an ontological and a phylogenetic sense.[1] Some caution is nevertheless in order. Those who insist upon this emphasis generally hold a materialist, i.e. nonvitalist, view of life—from which it follows that the difference between identifiably 'biological' structures and physical ones can only be a question of degree, and hence sufficiently elaborate physical (in the sense of nonbiological) structures should in principle be able to do everything biological ones can.

**MODULARITY**

The incredible and immensely complex biological structure of the brain, as well as the constant and rapid evolution of its interconnexions, may well prevent its detailed structure from ever being mapped out, but this should not cause undue despondency, for that knowledge may be largely irrelevant. In 1861 Broca believed that he had discovered a site of functional specialization (localization) within the brain [3]. From our present standpoint, almost a century and a half later, we can perceive that his discovery was steeped within the then prevailing view of the brain as an assembly of specialized units, as illustrated in figure 1. This view in fact persists to some extent to this day. Figure 2 shows Positron Emission Tomographs (PET) of the human brain while the subject was asked to carry out certain actions. In each case one or more different zones are active, Such experiments—admittedly carried out under rather artificial conditions—*appear* to show functional specialization, but the most that can actually be said is that the brain follows a strategy of assembling cells with common properties together. Such unification could equally well be, and sometimes is, achieved by distant neurons operating in synchrony. It would therefore be better to speak of *modularity* of the brain, the modules being groups of cells, while remaining aware that the basic unit of cortical organization is rather difficult to define. Furthermore, although there is evidence for multistage integration (e.g. in the visual cortex), there is no single cortical centre to which all other areas report exclusively, and if there were, to whom or what would it report to? Moreover, the end result of integration within the brain is starkly different from the integration within a digital computer: for example, when two numbers are multiplied together, knowledge of the individual numbers is lost when the result is computed, but in the brain's integration involved in, say, discerning a visual image ultimately identified as a "red flower", both the colour and the type of object must be preserved.

Despite their limitations, PET and other related

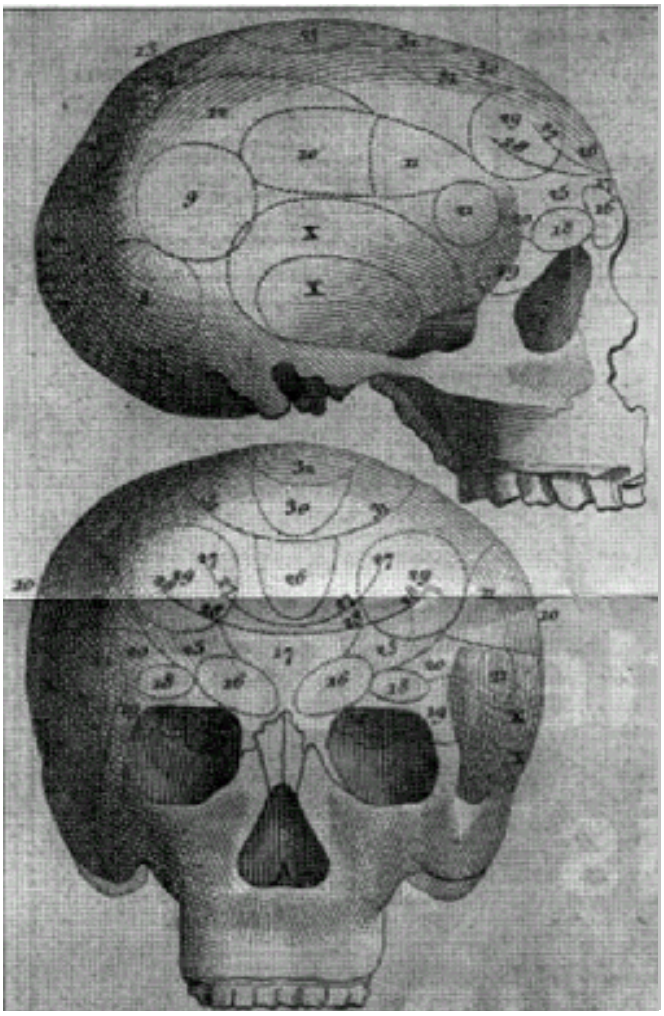

**Figure 1.** An illustration of attempts to localize psychical phenomena in the brain [4].

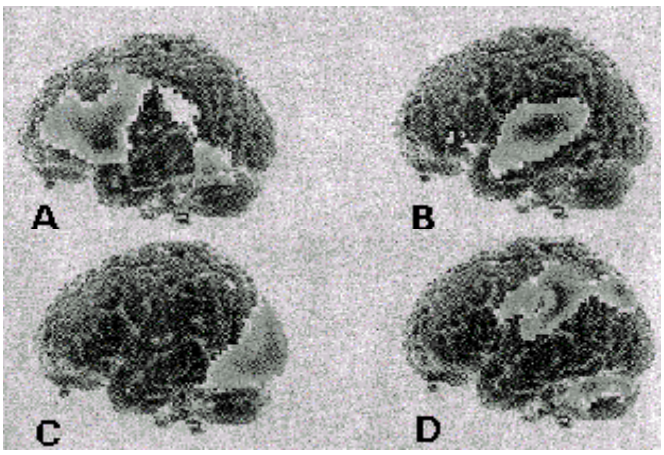


**Figure 2.** A Positron Emission Tomograph (PET) of the left half of a human brain, obtained by feeding the subject with radioactive sugar, whose emission (of positrons, that is, positive electrons) is detected with a certain spatial resolution. The intensity of the emission from a given region is proportional to the delivery of sugar to that region, and hence, in some sense, to the "intensity of thinking", since the sugar provides the necessary energy to active neurons, and neuronal activity is correlated with thought. The different images were taken after asking the subject (A) to think mutely of words, (B) to listen to words, (C) to look at written words, and (D) to feel a Braille script. Adapted from the *Neue Zürcher Zeitun*g, 18 March 1998, p. 65.

[1] In this context, it is worth nothing that the development of the brain in the embryo is very far from deterministic: there is no evidence that connexions between neurons are preprogrammed; at most there is a genetically-given algorithm to select favourite (in a certain sense) system connexions [2]. Thus the brains of genetically identical organisms would almost certainly be different from one another.

observations irreducibly point to an intimate relationship between the physical activity of the brain and the conscious experience of the individual. Not only does mental activity result in detectable physical changes in the brain, as illustrated in figure 2, but, conversely, minute physical changes introduced in the brain's pattern of activity, such as by drugs, or electric or magnetic fields, can engender changes in conscious experience, and even behaviour, although it must be emphasized that one cannot stimulate or destroy a given region of the brain and *reliably* produce only one type of behaviour in a subject [5].

**MECHANICAL ANALYSIS OF THOUGHT**

Despite the apparently tenuous parallel between electronic computing devices and living brains, the concepts of the digital computer and the neural circuit have at the very least rendered an important service by establishing the possibility of analyzing human experience in terms of mechanical activity. This demystification of thought can be made clear through considering some very simple model systems, in particular some simple examples of automata—networks of cells—of the kind first studied by Caianiello, Kauffman, Aleksander and others some decades ago. An *automaton* may be defined as a machine which processes information. Now whatever else the brain is or does, clearly it processes information, therefore it may be categorized as an automaton.

Consider an automaton consisting of just three cells, labelled A, B and C. Each cell has two input ports ("inputs") and one output port ("output"). For the sake of simplicity, but without any loss in generality, the inputs and outputs shall be considered to be able to have values of 0 or 1 only. This is scarcely a restriction, since practically any signal can be represented as a string of zeroes and ones, which in turn represent some physical quantity. For example, 1 could signify that a current (of water, or electricity) is flowing, and 0 the absence of current. The cells are joined together such that the output of each cell is connected to the input of the two others. The three cells are identical, and contain some internal machinery giving them the property that their output is 0, unless both inputs are 1.

The state of a cell is defined as its output at any given instant. It is straightforward to deduce that if all the cells start off in state 1 (i.e. with output 1), they will forever remain in that state. The same applies to an initial state of 0, but if any other starting combination is chosen, the automaton will move through sequences of different states (each of which is given by enumerating the states of the automaton's constitutive elements), until it finally enters the state in which all cells have state 0. This evolution can be captured in a diagram (figure 3), called the state diagram. The ensemble of these states and their interconnexions is called the *state structure*. In other words, state structure is the way the internal states change from state to state. There are just two stable states to which the system evolves, regardless of the starting values. That is, after the system has been running a certain time, one

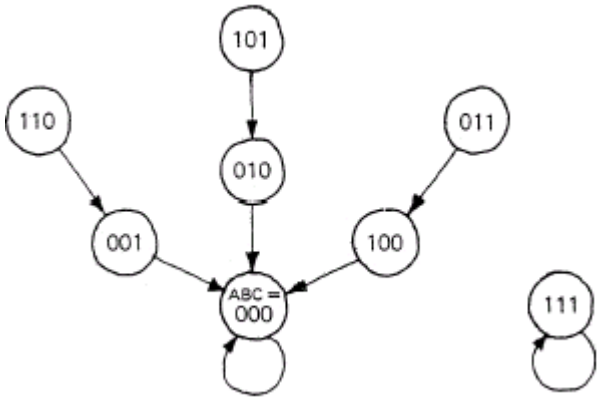


**Figure 3.** The state structure of a simple automaton [6] (see text).

can predict that it will be found only in one or the other of these two states.

If the functions of the elements are slightly changed, such that A becomes 1 whenever B is 1; B becomes 1 whenever C is 1; and C becomes 1 whenever A and B have the same value, the state structure becomes quite different. Although if the system happens to start in state {111} (the three digits denoting the states of A, B and C respectively) it will again forever remain there, starting off in any other state results in perpetual movement round a cycle of states, implying that the state in which the system is to be found at any given instant is quite unpredictable. If the elements are connected up differently, such that A's outputs go to itself and C; B's outputs go to itself, A and C; and C's output goes to B, the state structure remains identical, but if, keeping these changed interconnexions, the cells are restored to the functions they had in the first example, then the resulting state diagram turns out to be quite similar to the original one (figure 3). The reader is encouraged to verify these results and extend them, either using pencil and paper for simple networks, or with the help of a computer for those who are programming enthusiasts, which will make it practicable to look at the behaviour of larger and more elaborate networks.

The point of these illustrations is to demonstrate that state structure, which is no less real than physical structure, depends on both the physical structure (i.e. the interconnexions) and the functions of the constituent elements (cells): studying one without the other cannot lead to proper understanding. At the same time one notices that the state structure of a machine is only tenuously related to its physical structure. Therefore, although much has been made of the need to consider the biology of the brain, especially the biological features of its construction, at a certain level of abstraction this is *not* actually very important. Automata with very different physical structures can have identical state structures (think of a computer memory and a mammalian brain, physically very different, yet both can remember), and similar physical structures can have very different state structures, as has just been shown. It follows that it is futile to devote a great deal of time to the study of physical structure: careful charting of all the connexions of the nerve cells

in the brain will not reveal what the brain is doing.

The concept of *state activity* refers to the way that the state of a system changes from one time to another. Notice the kinetic emphasis: thinking should be thought of in terms of a *process*, rather than something static; thoughts are to be identified with state trajectories, i.e. trajectories among its states [6].

There is one further point which should be made in connexion with these networks of cells. By replacing one of the inputs, say from B into A, in the original network of three cells (in which each cell has two inputs from the other two cells) with an input from the external environment, one obtains two new state structures, depending on whether the input from the environment is 1 or 0, demonstrating the important point that cellular networks have state structures closely related to the nature of the information impinging on the network from without. In other words, state structure is a kind of reflexion of the environment—one that makes sense of the world to which the network is exposed.

The interaction of the automaton with its environment can be further extended by giving B (for example) an output to the environment: the combination of automaton and environment creates another autonomous (i.e. with neither inputs nor outputs) automaton. It might furthermore be reasonable to consider the environment as a probabilistic automaton, i.e. its output will be 1 with a certain probability.

One can continue in this vein, constructing more and more elaborate networks, with supervisory and self-supervisory features, which show attributes such as learning, memory, awareness, dreaming and so on, attributes which begin to more closely resemble those we associate with our own minds [6]. One should emphasize, however, that a deep explanation of thought has not been thereby provided: the total number of states of the brain is so immense, and its state structures correspondingly complex, that it has scarcely trivialized thought by associating it with state structures. At this stage one should merely note that it appears to be possible to explain rather abstruse philosophical concepts, including will and emotions, in terms of simple machinery rather than mystical concepts,[2] without implying that such machinery actually exists in the brain.

**REDUCTIONISM**

The mechanistic arguments developed in the previous section would appear to support the reductionist view ahead of its rivals. One must, however, be careful not to fall into the trap of "nothing-buttery", to use a phrase coined by Donald MacKay [7]: it would be unwarranted to view the brain as *nothing but* the mindless motion of molecules.[3] Despite the large amounts of neurophysiological information now available, the anticipated equation which would reduce consciousness to matter has not in fact been given. The reductionist programme is at best a matter of *correlation*; that is, when there is consciousness, there is a certain kind of electrical and chemical activity in the brain, and when there is not consciousness, there is not that certain kind, but electrical and chemical activity of another kind, or none at all (cf. figure 2). More cannot be deduced from the available data. No actual *identity* has ever been shown to exist between a thought, an awareness, a concept, an intention, a meaning, or any other kind of mental happening, and a particular group of material happenings, i.e. neural events in a brain. The reduction is thus, at least until now, evidentially ungrounded.

By way of illustration, if a certain number, say 18 769, is represented in the register of a digital computer, certain things must be true about the physical activity of that computer which would not be true if that number were –137, but there are no grounds for asserting *identity* between the physical states and the symbol. Moreover, since computers are designed to tolerate variations in power supply and changing characteristics of ageing components, quite a wide range of physical states may in fact have the same symbolic significance.

Let us for a moment assume that the form of your physiological brain processes indeed determines the content of your conscious experience of thinking, feeling and understanding, and that you write a description of your experience of some event down on one side of a table, and call it the 'I-story'. A superphysiologist who knows the complete pattern of correlations will be able to make a corresponding entry for his or her description of your brain in neurophysiological terms, and can write it down on the other side of the table as the cerebral correlate of your own entry. The 'I-story' about your immediate experience is thus *correlated* with the brain story about a physical structure and the physiological interactions supposedly taking place in it. It is by no means necessary to insist that the correlation is one to one; the same conscious experience could result from several different patterns of neural firing, and vice versa (as illustrated by the analysis of state structures of cellular automata), and not every human brain activity has a correlate in conscious experience; consider for example the neural mechanisms that regulate normal breathing.

Hence from the one, we may infer the other, but to suppose that this tight correlation between physical and mental states justifies attributing activities in one story to entities in the other is simply wrong. To see this perhaps more clearly, suppose a digital computer has been set up to go through a specific program, say one that can solve quadratic equations. At any point in its operation there is

[2] Such as the dualist view espoused by Popper, Eccles and others, according to which brain and mind are separate, distinct realities (it is sometimes asserted that mind *emerges* from brain processes): one of the main arguments against the dualist view is simply that no valid evidence demands it, rather than that it can be decisively refuted.

[3] Or, as Erwin Schrödinger put it, "...sind molekelstosse nur."

a strict correlation between the mathematical or logical function being executed and the physical activity of its transistors and other elements. Someone who knows the machine in detail can in principle derive the 'machine story' corresponding to every entry in the functional story. But this in no way implies that the story about the machine and its mechanistic working *states the same facts* as the story about the mathematical and logical functions embodied in those workings, *for the two stories are in different categories*. The fact that a quadratic equation with two roots is embodied in a piece of electronic hardware in no way implies that the hardware, at any level of description as hardware, 'has two roots'. The notion makes no sense, even though the existence of two roots has perfectly well defined hardware implications [7].

**EMBODIMENT AND AGENCY**

Thinking, our conscious experience, is thus *embodied* in our brain activity, but that is far from saying that it is the same thing. The evidence shows only that we are embodied in our biological structures, and the things that we do as (conscious) cognitive agents—desiring, planning, observing, understanding, acting, etc.—are mostly not meaningfully defined if attributed to biological structures, any more than to nonbiological ones.

Whenever one of these activities is referred to, the subject of the verb is generally 'I': we are bearing personal witness as truthfully and explicitly as we can to the immediate data of our conscious experience, data which include the experiences of seeing our limbs move [8], longing for a cup of tea, etc. It is thus as *agents* that we think and understand, and have first-hand knowledge of what this means, and nowhere do we find any meaningful basis grounded on solid evidence for attributing the experiences of conscious agency to the physical (or biological) structures in which they are embodied.

An agent has three essential attributes: (i) a *repertoire* from which alternative actions can be selected; (ii) an *evaluator* which assigns values to different states of affairs according to either given or self-set criteria; and (iii) a *selector* which selects actions increasing positive evaluation, and diminishing deleterious evaluation. Artificial agents such as autopilots in aeroplanes bearing these attributes can and have been constructed. But do they understand? Undoubtedly they can *mis*understand—a machine assisting a mechanic may respond to a shout of "Wait!" by placing an object on a weighing machine and calling out its weight—and, if the notion of misunderstanding makes sense to us, it would seem illogical to deny that they can understand too. *For any agent, the meaning of an item of information is ultimately realized in terms of the contribution it makes to the agent's total state of conditional readiness for action and the planning of action* [9], i.e. its conditional repertoire of action, and there seems to be no reason why an artificial agent cannot be thus endowed, just as much as a living biological one. It should therefore be acceptable that it makes sense to attribute understanding to agents, whether real or artificial, but that is not at all the same as saying that a brain, or a computer, can understand: agency is embodied in the brain or computer.

**IS THOUGHT ALGORITHMIC?**

Reductionism might still be rescuable if it could be shown that the thought processes undergone by the agent follow rules, i.e. are algorithmic. Consider the chess position shown in figure 4. White is to play and draw—an easy enough problem for a human player, but when the computer Deep Thought was presented with this situation,

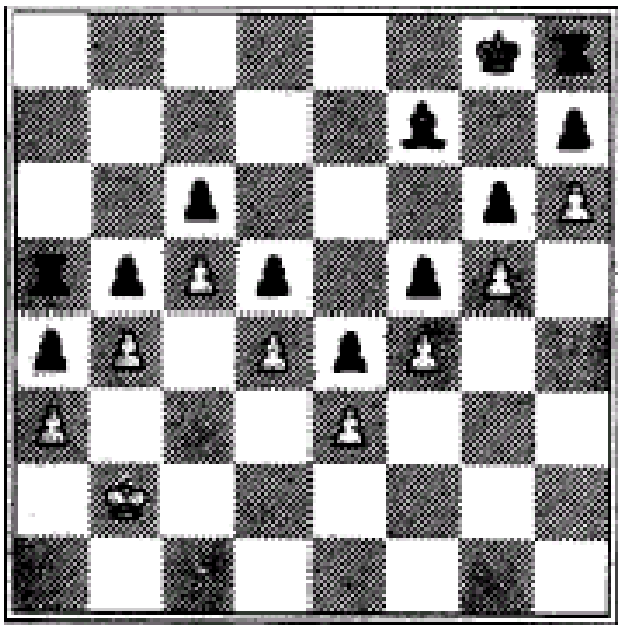

**Figure 4.** White is to play and draw [10]. See text.

it took the castle [10], thus indicating that computers have not the remotest *understanding* of the game. There appears to be something in understanding which is not the same as computation (i.e. rule following): rather, the two are totally different kinds of issues; in other words, they are in different categories.

The difference crops up all the time in mathematical reasoning. Penrose has given a good example of non-computability in formal terms, based on a theorem due to Berger, which proves that there is no uniform computational procedure for deciding whether a given nonperiodic set of polyominoes (generalized dominoes, made up of any number of squares) will tile the Euclidean plane [10]. Take a finite set of polyominoes numbered $S_0$, $S_1$ etc., even subscripts corresponding to even numbers of squares, and allow time to evolve such that at each step the next subscripted polyomino or set of polyominoes is taken, except that a number is skipped if the corresponding set does not tile the plane. The state of the universe at any epoch $i$ is given by the set of polyominoes $S_i$. The problem to be solved (by computation or otherwise) is: at any given epoch can the entire Euclidean plane be covered using the shapes in the given set?

The precise evolution rule is completely deterministic (since the set of polyominoes is given beforehand) but not computable, i.e. it cannot be simulated with a

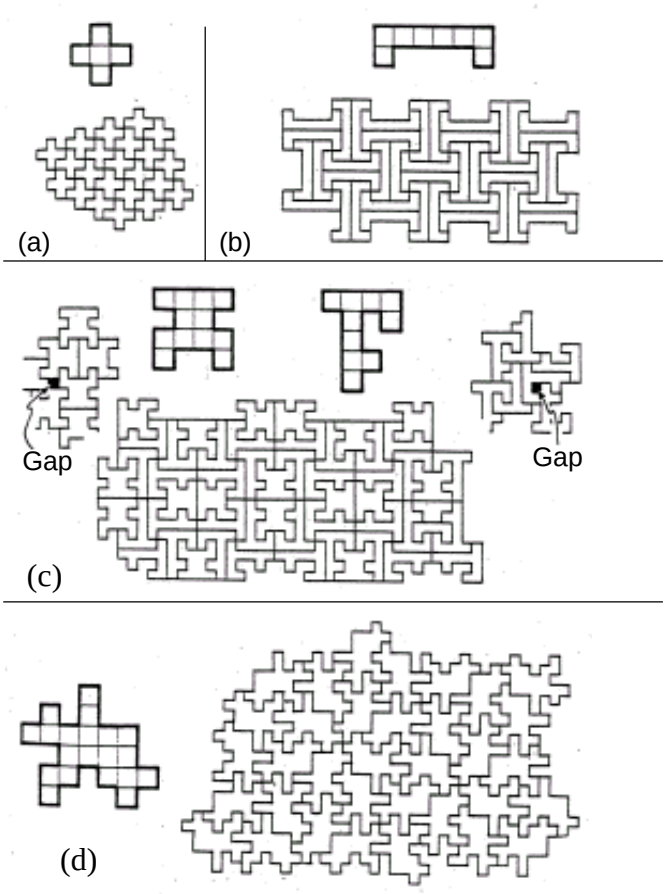


**Figure 5.** Polyominoes [10]. See text.

computer, since there is no uniform computational procedure. Figure 5 may make this clearer. Shape (a) does tile the plane, shapes (b) and (c) together will tile the plane, but neither does so on its own, and shape (d) does tile the plane but not in any repeating way. Whatever procedure one specifies, there is always some other set which gets outside that procedure, and it is in this sense that the universe under consideration is not computable.

Harkleroad gives another example of noncomputability [11]. Let the programs running on a (finite) computer be codenumbered as follows: $P_n$ is the program with code number $n$, and the function specified by $P_n$ is denoted $f_n$ (whose domain and range are taken to be subsets of $Z^+$, the set of positive integers). Now consider the function $g$ (also with domain $Z^+$) defined by:

$$g(n) = \begin{cases} 2 & \text{if } f_n(n) = 1 \\ 1 & \text{otherwise.} \end{cases} \quad (1)$$

Now $g$ cannot be the function $f_M$, regardless of the number $M$, because $g$ and $f_M$ behave differently when applied to $M$. Namely, either $g(M) = 2$, in which case $f_M(M)) = 1 \neq 2$, or $g(M) = 1$, in which case $f_M(M) = 1$. Thus no program specifies $g$. But why cannot the following be formalized as a legitimate program?

1. Given input $n$, determine the program $P_n$.
2. Then feed $P_n$ the input $n$.
3. If $P_n$ returns 1, then give 2 as output for $g$, else give 1 as output.

The snag occurs in the case when $P_n$ doesn't return any output—as long as there isn't any (and the computer cannot be programmed to find out whether there will be, since $g$ can't be programmed), one is simply kept waiting, for what might turn out to be infinite time!

These examples illustrating the notion of noncomputability are mathematical, but that does not restrict the domain of applicability of the argument. It has been shown that at least some mathematical understanding is noncomputable (nonalgorithmic); but mathematical understanding is just a subset of the totality of human understanding, without being clearly demarcated from it. Since there are no grounds for clearly demarcating human understanding from human consciousness, which in turn cannot be clearly demarcated from animal consciousness, the nonalgorithmic quality should apply to all consciousness.

**INDETERMINACY**

Ultimately one must consider the detailed nature of the processes determining state transitions within the brain. In a perceptive comment, MacKay remarks that "state transitions are often determined not by consulting rules but by what amount to local physical experiments, usually with a stochastic ingredient which can give rise to spontaneous (though not statistically nonsensical) turns of events" [12]. When that remark was first made, very little was known about processes within neurons, but since then a great deal (although, one can be sure, far from everything) has been discovered. Hameroff and Penrose have argued in favour of the microtubules within neurons being sites of nonalgorithmic quantum computing [13]; an essential feature of quantum mechanical objects is that they can exist in a superposition of states, and the manipulation of these states prior to reduction ('collapse') could provide a physical basis for nonalgorithmic reasoning.

Quantum indeterminacy has an interesting correlate in logical indeterminacy. The Principle of Logical Indeterminacy states that there does not exist any specification of our cognitive mechanism unknown to us with a unique and unequivocal claim to the assent of everyone if only they knew it [14]. The specification of our cognitive mechanism at a certain time should have an unconditional claim to the assent of the superphysiologist who obtains it (see above), but *we* would have a strong logical reason for *not* assenting, since it is accurate only in relation to our present state. If we were to believe it, then the specification would no longer be accurate. There can be no specification which is equally accurate whether or not we assent; hence some aspects of the state of our cognitive mechanism are indeterminate for us, and a prediction of the state cannot logically be labelled 'true' or 'false'.

There are several corollaries flowing from this concept of logical indeterminacy, one of which being that agents can have free will, but not the brain or computer in which agency may be embodied. Thus, despite the intimate two way relationship between the physical activity of the brain

and the conscious activity of the individual, a human being has an individual identity which is no more tied down to a particular brain structure at a particular epoch, any more than the state of an interacting dynamical system is tied down to knowledge of all the positions and momenta of its constituent particles at a particular epoch. Our identities are defined by the unique progression of personal acts and decisions made throughout life.

**THE RÔLE OF PROPRIOCEPTION**

These 'I'-dentities are more closely linked to our bodies, rather than merely to our brains, than it is currently fashionable to admit. The immediate data of conscious experience is acquired in the first place by proprioception, that is, the sense of movement and position (recalling Sherrington's definition of proprioceptors as 'sensory organs stimulated by actions of the body' [15]), and especially kinesthesia, the sense of movement through muscular effort. There are excellent reasons for ascribing to proprioception the primacy among the senses normally reserved for vision. Sensitivity to movement is basic and paramount for an animate creature. Its environment is constantly changing, in ways that are too complex and demanding responses too complex to be centrally programmable. Even a beetle must constantly adjust kinetically to minute declivities and tiny grains of sand which it dislodges. Indisputably, its medium influences what an animal can do and what it actually does. Proprioception provides information about both the body and its surroundings, information which vision alone is unable to provide. We do not merely look and see, *pace* David Hume, Auguste Comte *et al.* (cf. [8]). Beginning with external proprioceptors, proprioception evolved from tactility to kinesthesia and was internalized along the way, which allowed better discrimination between movement of the body and movement generated by external events, as well as being less vulnerable to damage, and following the ontogeny/phylogeny parallel alluded to earlier, the argument applies as much to the embryological development of the brain of an individual as to the evolution of animals.

The kinetic spontaneity associated with dealing with the environment is so characteristic of animation—Alexander Bain long ago pointed out the contrast with the inanimacy of a merely falling stone [16]—and proprioception is so essential for enabling that response, endowing matter, now animate, with a sense of agency, with 'I can', that it seems almost inescapable to concede that proprioception arose with the appearance of animate forms and is inextricably associated with a sense of self—we perceive the qualia of our own movement. Proprioception links corporeal consciousness to the level of actual life, i.e. to movement and *experiences of moving*; the relationship is moreover two-way, for these experiences affect animal form, and movement itself is conditioned by animal shape and pattern. Hence by specifying how consciousness takes into account the actual lives of individual animate forms—creatures—we have approached an answer to the central question, as it is usually considered to be, of *how* consciousness 'arises' in matter. We are not constrained to awkwardly aver, with richly metaphysical overtones, that consciousness somehow arises in matter, instead we affirm that it arises in animate forms; it inheres, or is rooted, in the animate; and we insist that the question, '*how* is it embodied?' makes no sense, for it implies a mixing of categories, nor does Churchland's insistance that "the important point about the standard evolutionary story is that the human species *and all its features* are the wholly physical outcome of a purely physical process" [17] (emphasis added). Kinetic cognitional activities constitute a corporeal consciousness which is as vital a biological faculty as those more familiarly mentioned as attributes of life—self-replication, etc. Hence it can truly be asserted that we as individuals are embodied, and our self-awareness, our consciousness (it might be noted in passing that it has never been shown that unconsciousness preceded consciousness) is rooted therein and is not some metaphysical "higher order" function.

As emphasized by Maxine Sheets-Johnstone [18], creatures know themselves in ways that are fundamentally and quintessentially consistent with the bodies that they are. Even a bacterium can apparently sense both the environment (via the proton motive force in its body, for example) and itself with respect to its environment. To ascribe consciousness uniquely to humans seems unwarrantedly arbitrary, and to assert that it arose with language, considered to be a uniquely human attribute, merely begs the question how language arose. Proprioception thus stands out as an epistemological gateway, enabling corporeal consciousness, which subsequently expanded into a sense of self. This argument may be compared with Maximov's interesting theory of how colour vision arose: creatures living in the sea close to the shore had to develop the ability to maintain constancy of perception of prey whose (monochrome) image was superimposed upon the constantly moving shadows of the waves. The expansion of neurological capabilities which this ability presupposes would have been an essential prerequisite to the development of colour vision from monochrome [19].

**CONCLUDING REMARKS**

While many of the attributes of human brains can be plausibly accounted for by conceptually simple mechanical models, the crucial reductionist link of *identity* between thoughts and mechanical events has not so far been even remotely established. At best, one can demonstrate correlation. It is wiser to assert that thinking is *embodied* in brain activity. It is as embodied *agents* that we think and understand; and there seems to be reason why artificial agents sharing these attributes could not be constructed.

Embodiment is one obstacle to reductionism; the apparently irrefutably non-algorithmic nature of at least some thought processes is another. Furthermore, logical

indeterminacy ensures that agents, but not the hardware in which they are embodied, can have free will.

Indubitably, mental activity is represented by detectable physical changes in the brain, and minute physical changes introduced in the brain's pattern of activity can engender changes in conscious experience. Curiously, a fairly obvious corollary seems to have hitherto received scant attention from researchers: namely the possibility that conceptual input to our minds—the ideas, images etc. with which we are daily flooded—may play a rôle in engendering neurodegenerative disease, i.e. physical changes in our brains. If this is so, then we should pay as much attention to our conceptual environment as our physical environment currently receives.

**REFERENCES**


1. Goodacre, R., Kell, D.B. and Bianchi, G. Nature (Lond.) **359** (1992) 594.
2. Érdi, P. and Barna, Gy. Biol. Cybernetics **51** (1984) 93–101.
3. Broca, P.P. Bull. Soc. Anthrop. **2** (1861) 235–238.
4. Bojanus, L. Phil. Mag. **14** (1802) 77–84.
5. Valenstein, E., in *Modifying Man*: *Implications and Ethics*, ed. C.W. Ellison. Washington, D.C.: University Press of America, 1978.
6. Aleksander, I. *The Human Machine*. St Saphorin: Georgi Publishing Co., 1977.
7. MacKay, D.M. *Human Science and Human Dignity*. Sevenoaks: Hodder and Stoughton, 1979.
8. Graziano, M.S.A., Cooke, D.F. and Taylor, C.S.R. Science **290** (2000) 1782.
9. MacKay, D.M. Sci. Christian Belief **1** (1989) 27–39.
10. Penrose, R. J. Consciousness Studies **1** (1994) 241–249.
11. Harkleroad, L. College Math. J. **27** (1996) 37–42.
12. MacKay, D.M. Synthèse **9** (1954) 182–198.
13. Hameroff, S. and Penrose, R., in *Towards a Science of Consciousness*, eds S. Hameroff, A. Kaszniak and A. Scott Cambridge, Mass.: MIT Press, 1996.
14. MacKay, D.M. *Behind the Eye*, ed. V. Mackay. Oxford: Blackwell, 1991.
15. Sherrington, C.S. *The integrative action of the nervous system*. New York: Charles Scribner's Sons, 1906.
16. Bain, A. *The Senses and the Intellect*. London: Parker, 1855.
17. Churchland, P.M. *Matter and Consciousness*. Cambridge, Mass.: Bradford/MIT Press, 1984.
18. Sheets-Johnstone, M. J. Consciousness Studies **5** (1998) 260–294.
19. Maximov, V., in Proc. 2nd Conv. Georgian Physiol., Tbilisi (2000), pp. 119–121.