The unsolvability of the mind-body problem liberates the will

Jan Scheffel
Professor, Fusion Plasma Physics, KTH Royal Inst. of Technology, Stockholm, Sweden

Abstract

The mind-body problem is analyzed in a physicalist perspective. By combining the concepts of emergence and algorithmic information theory in a thought experiment employing a basic nonlinear process, it is argued that epistemically strongly emergent properties may develop in a physical system. A comparison with the significantly more complex neural network of the brain shows that also consciousness is epistemically emergent in a strong sense. Thus reductionist understanding of consciousness appears not possible; the mind-body problem does not have a reductionist solution. The ontologically emergent character of consciousness is then identified from a combinatorial analysis relating to system limits set by quantum mechanics, implying that consciousness is fundamentally irreducible to low-level phenomena. In the perspective of a modified definition of free will, the character of the physical interactions of the brain's neural system is subsequently studied. As an ontologically open system, it is asserted that its future states are undeterminable in principle. We argue that this leads to freedom of the will.

1 Introduction

Understanding consciousness is a central problem in philosophy. The literature produced through the centuries, relating to the 'mind-body' - problem, is also huge. A subset of over 3500 articles has been collected by Chalmers (2016). An apparent contradiction lies in the fact that while we normally seek scientific understanding from a reductionist perspective, in which the whole is understood from its constituents, consciousness has for millions of years evolved into an extremely complex system with advanced high-level properties.

The theoretical difficulties we have faced strongly suggest that fundamentally new ideas are needed for the mind-body problem to reach its resolution. In this work it is argued that emergence, combined with results from algorithmic information theory, is such an idea. The meaning of these concepts will shortly be discussed; we may here briefly state that emergence relates to complex systems with characteristics that are difficult or impossible to reduce to the parts of the systems and algorithmic information theory concerns relationships between information and computing capacity. We reach the conclusion that the mind is epistemically strongly emergent, which by definition implies that the mind-body problem cannot be solved reductionistically. Reductionistic understanding of the subjective aspects of consciousness, like introspection and qualia, is therefore not possible. The concept of 'explanatory gap' (Levine, 1983) is thus justified.

McGinn, in his influential work "Can we solve the mind-body problem?" (McGinn, 1989) also concludes that the mind cannot be understood, but on other grounds. He focuses on the ability to understand phenomenal consciousness (Block, 1995), and find that we humans, because of "cognitive closure" are not able to solve this 'hard problem of consciousness' (Chalmers, 1995). With the reservation "the type of mind that can solve it is going to be very different from our" McGinn does not fully exclude that consciousness can be given some kind of explanation, an optimism not supported in this work.

But how is freedom of the will related to the mind-body problem? We have devised a thought experiment which features a process that can be shown to be strongly emergent in the epistemic sense. The close relation to neurological functions of consciousness leads to the conclusion that also consciousness is epistemically strongly emergent. Furthermore we will argue that consciousness, due to its particular complexity considered as a global system, is ontologically emergent. Chalmers (2006) finds, on intuitive rather than on formal grounds, that the mind is 'strongly emergent'; a term used in the same meaning as 'ontologically emergent'.

The ontologically emergent character of consciousness resolves the contradiction we have long been facing for the freedom of the will; on the one hand, our volitional actions should occur deterministically quite as we wish to see them executed, but on the other hand determinism should not force our thoughts and intentions upon us. This explains the seemingly strange choice of title for this work. Of crucial importance is the use of a reformulation of the common definition 'ability to do otherwise' to a scientifically testable definition of free will. We here also touch upon the somewhat neglected circumstance that if we indeed could understand consciousness, it follows that the will cannot be free.

Definitions are important in this work. There are at least two reasons for this. The first is that several aspects of the concepts of consciousness and freedom of will often are used in different ways by different philosophers, neuroscientists and others. Furthermore, it is sometimes unclear which definition is assumed. This may be understandable on the basis of that consciousness, not least semantically, is an elusive concept. The problem is rooted in its unique character, causing attempts for a definition to contain circular elements of some kind. The influential early characterisation of Locke (1690) "consciousness is the perception of what passes in a man's own mind" suffers from reference to the subjective term "perception". Nagel's (1974) characterisation "there is something that it is like to be that organism - something it is like for the organism" has gained popularity, although "is like" refers back to the subject itself, that is to consciousness. A more exhaustive and recent discussion of possible criteria for and meanings of conscious states can be found in Van Gulick (2014). However, either of the above formulations sufficiently catches the subjective components of consciousness that are referred to in this work, thus we here consider phenomenal consciousness. When we discuss other key concepts, attempts will be made to render the treatment more precise, in some cases using formalisations from physics and mathematics.

The second reason for the need for clear definitions is simply that binding arguments requires precision (Carnap, 1950). The consequence of such specifications may of course be that the definitions of some philosophers are excluded; the results of this article should be seen in this perspective.

We begin by discussing what requirements must be placed on a solution of the mind-body problem. It is then argued that such a solution cannot be found. The core of the argument is that the *epistemically* strongly emergent nature of consciousness precludes understanding of it in a reductionist sense. In the subsequent section we show that the *ontologically* emergent character of consciousness dissolves the deterministic contradiction we have been facing for the freedom of the will. Finally follows discussion and conclusion.

2 What is required of a solution to the mind-body problem?

The goal of the mind-body problem research is to find a theory that explains the relationship between mental and physical states and processes. The sub-problem which by far has attracted the most interest concerns the question how consciousness can be understood. We may initially ask the question: what is required from an adequate theory?

Chaitin (1987) has clarified the meaning of the necessary requirement that a theory must be inherently less complex than what it describes; in his terminology it must to some extent be 'algorithmically compressible' in relation to what it should explain. Let us illustrate this by an example. A relationship $y = f(x)$ has been established to explain a phenomenon, but the precise dependence is not known. A series of experiments that generate N data points (x_i,y_i); i:1…N has thus been performed. Clearly, a polynomial $Y(x)$ of degree N-1 (N coefficients) can always be fitted to be drawn through all the data points in an xy-diagram. Is it a theory? The answer is no, for the simple reason that $Y(x)$ does not explain anything; it is always possible to draw a polynomial of degree N-1 exactly through N data points. Had we instead adapted a polynomial of lower degree through all the points, say a second order polynomial through 10 data points, then we would have a theory worthy of the name; it predicts more than it must. That it is algorithmically compressible means that it can be formulated using fewer bits of binary information than for $Y(x)$. Simply put: *a proper theory must be simpler than the phenomenon it describes, otherwise it does not explain anything*.

We will in this work make use of the discrete logistic equation for comparisons with the neurological processes that form the basis of consciousness. This equation can be formulated as the discrete recursive relation $x_{n+1} = \lambda x_n(1-x_n)$, n:0…n_{max}, where the positive integer n_{max} can be chosen freely. The discrete logistic equation then iteratively generates new numbers x_{n+1} for increasing values of numbers n. The parameter λ and the start value x_0 must first be selected. We can now ask: is there an explicit theory for the value x_{n+1}, that is is there a function $u(k)$ which satisfies the relation $x_k = u(k)$, being algorithmically compressible as compared to repeated use of the iterative relationship $x_{n+1} = \lambda x_n(1-x_n)$? Of course we can form $x_1 = \lambda x_0(1-x_0)$, $x_2 = \lambda x_1(1-x_1) = \lambda\lambda x_0(1-x_0)(1-\lambda x_0(1-x_0))$ and so on. This latter route is however not feasible; for large n we will find that x_{n+1} becomes extremely complex; this way of searching $u(k)$ does not result in a valid theory. Alternatively formulated: the binary bits needed to represent these terms is at least of the same order as the bits representing the numbers x_1, x_2, x_3... themselves! Unfortunately, it can be shown that the question must be answered in the negative; no matter how we try it is not possible (except for a very few values of λ) to derive a theory, that is a compact, explicit expression for $u(k)$.

The cause of the problem is that the discrete logistic equation is a *nonlinear recursive* equation. Let us, for a moment, instead consider the simpler *linear recursive* equation $x_{n+1} = A+\lambda x_n$, n:0…n_{max}, where A is a constant, for which the general term x_k can be derived in *explicit* form simply as $x_k = x_0\lambda^k+A(1-\lambda^k)/(1-\lambda)$ for $\lambda \neq 1$ and as $x_k = x_0+Ak$ when $\lambda = 1$. The solution is expressed using only a few mathematical symbols; it is thus algorithmically compressible (may be represented by fewer digital bits of information) as compared to the solution x_k obtained iteratively by forming $x_1 = A+\lambda x_0$, $x_2 = A+\lambda x_1 = A+\lambda(A+\lambda x_0)$, $x_3 = A+\lambda x_2 = A+\lambda(A+\lambda(A+\lambda x_0))$ and so on. This explicit solution was analytically available because of the low complexity involved in the solution of linear equations as compared to nonlinear. Furthermore, the solution for x_k is derived mathematically by using well-known axioms and theorems; consequently we can *theoretically explain* the values x_k for the linear recursive equation.

We have here employed examples from mathematics, but the reasoning applies generally when we seek any kind of formal explanation or theory for a phenomenon. As a result, a theory cannot explain consciousness if it relates to systems of the same level of complexity (like other minds). Understanding is only reached from theories that are less complex than consciousness itself and relate to already established knowledge; in other words they should be algorithmically compressible in relation to consciousness.

3 Emergence stands in the way

The emergent character of consciousness is still debated in the philosophical literature (Kim 1999, Kim 2006, Chalmers 2006). We will here argue that consciousness is both epistemically strongly emergent and ontologically emergent. Consciousness thus has features that are not reducible to the properties of its components. The standard expression 'not reducible to' expresses that the characteristics of the low-level components of the phenomenon are insufficient to determine high-level properties. By 'low-level' and 'high-level' we refer to the parts of and integrated wholes of a system or phenomenon, respectively. This is of course central to the mind-body problem since it settles the issue of the 'explanatory gap' (Levine, 1983); an unbridgeable gap exists between the theories we can formulate on the basis of the basic physiology of the brain and the subjective, cognitive function of consciousness. A consequence is that behaviour of consciousness is unpredictable, a relationship being of importance as we later approach the problem of freedom of will. First we turn to investigate the emergent character of consciousness.

3.1 Definition of epistemical and ontological emergence

For this we define *epistemically strong emergence* in the following way: *A high-level phenomenon is epistemically strongly emergent with respect to phenomena on low-level if the latter constitute the basis for the high-level phenomenon and if the theories that describe the low-level phenomena cannot predict behaviour at high-level.*

Ontological emergence, in turn, can be defined by replacing "*if the theories that describe the low-level phenomena cannot predict behaviour at high-level*" with "*if its behaviour is not reducible to processes at low-level*". It may be argued that any phenomen or behaviour in a physicalist world would be reducible to low-level properties since we assume that the physical is all there is. But this is not what is meant by reduction; supervenience does not imply reducibility. An ontologically irreducible phenomenon, if it exists, would not be determined by its low-level-properties or behaviour; it would not follow a statistical or law-like behaviour in relation to its low-level components. Loosely formulated it can be said that its behaviour comes as a surprise to nature. This distinction is crucial and we will indeed see that even if determinism holds, there are systems where extreme complexity can, in an ontological sense, ’shield’ the dynamics of a high-level phenomenon from that of its associated low-level phenomena. An important consequence is that these systems are uncontrollable in principle. It may here also be remarked that there is no need to distinguish between 'strong' and 'weak' ontological emergence.

The requirements for ontological emergence are indeed harder to satisfy than those for epistemic emergence; the former relates to intrinsic properties of the system rather than to knowledge about and theories for the system. The term *epistemically weakly emergent*, in turn, is frequently used for systems that can be simulated on a computer but otherwise would be characterized as epistemically strongly emergent. This definition will be adopted here as well. Emergence is naturally related to complexity. The brain features about 100 billion nerve cells (neurons), each connected to thousands of other nerve cells via dendrites. A model of the mind must be able to handle a corresponding complexity. As we have just discussed, algorithmic information theory implies that 'models' or 'theories' that cannot be algorithmically compressed to a complexity lower than that of the data they describe do not measure up. It is however not entirely clear how emergent properties arise in complex systems. It would be of great help if we could actually point to a relevant example. Our approach will thus be to, using a thought experiment, provide an example of a system with strongly emergent properties, related to neural processes of the brain but with lower complexity. The epistemically strongly emergent character of consciousness will then follow.

3.2 The Jumping robot

Our thought experiment is the following. Let us imagine a number of robots that are deployed on an isolated island. All robots are designed in the same way. They are programmed to be able to freely walk around the island and perform certain tasks. The robots can communicate with each other and are also instructed to carry out their duties as effectively as possible. If a robot becomes more efficient by performing a certain action, it should 'memorize' it and 'teach' the other robots the same skill. Let us concentrate on the behaviour of one of these robots and call the thought experiment 'the Jumping robot'.

In order to support the robots to move about freely, their movement patterns are partially determined by the discrete logistic equation just described. The iterative equation $x_{n+1} = \lambda x_n(1-x_n)$ generates new numbers x_{n+1} in the interval [0,1] when x_0 (also in the range [0,1]), and λ are set. These numbers affect how the robot should coordinate its joints, muscles and body parts, but the robot is programmed only to use information leading to safe motion without falling. Let us put $\lambda = 4$. It can then be shown mathematically that, for almost any choice of x_0, a chaotic sequence in the interval [0,1] is generated already for moderately large n.

A chaotic sequence is characterised by that it is not algorithmically compressible (Chaitin, 1987); there is no theory that can predict the value of x_k for large k. If we consider consecutive x_k, x_{k+1}, x_{k+2} and so on, these numbers will seem completely random. But the interesting and important fact is that they actually are deterministic; each number in the sequence is unambiguously defined by the former and so on in a long chain. Let us summarize: in the general case, there exists no algorithmically compressible explicit expression $x_k = u(k)$ for the discrete logistic equation and for the case $\lambda = 4$, there is neither any empirical opportunity to establish a law for the generated values.

Now assume that it would be of great value if the robots could perform jumps without falling. An attempt is thus made to learn a robot this behaviour. From a large number of x_0 values different sequences of numbers are generated, using the discrete logistic equation, in the hope that one of these sequences would correspond to movements which when combined result in the robot performing a jump. We ignore here that the procedure is obviously cumbersome; the complexity is partly caused by the fact that the robot consists of a large amount of joints, muscles and other bodyparts that should be coordinated, partly by the ignorance as to what movements the robot would need to perform for a successful jump and partly by that the discrete logistic equation does not allow control of the movements. After numerous unsuccessful attempts the task is thus given up; the robots cannot be 'taught' to jump.

Instead now initiate robots with random x_0 and leave them to themselves for some time on the island, after which we return. To our surprise, we now find that several of the robots make their way not only by walking, but also by jumping over obstacles. They have thus developed an *emergent property or behaviour*. We cannot explain how one or more of the robots aquired the property to be able to jump; no theory is to be found. As we saw, we could neither simulate the behaviour. If so, this would have been an example of epistemically weak emergence. The robot's behaviour is thus *epistemically strongly emergent*. A main point here is that the emergent ability to jump *per se* is both fully comprehensible to us as well as fully plausible in the sense that we can imagine that a certain sequential use of joints, muscles and bodyparts may indeed accomplish this behaviour, at the same time realizing that some kind of chance or evolution beyond our modelling capacity was required in the light of the complexity involved. There is no magic involved in the process, rather the behaviour is similar to that of random mutations in the genome of an individual organism, producing improved characteristics. The behaviour, however, does not appear to be ontologically emergent since we know that the robot's capacity to jump really is reducible to its finite number of parts. Similar conclusions about the emergent properties of nonlinear systems have been reached by other authors (Silberstein and McGeever, 1999).

3.3 The emergent character of consciousness

What then is the relevance of this epistemically strongly emergent system for the mind-body problem? It could be argued that we can make detailed studies of a jumping robot, simply ignoring how it reached its emergent state, in order to understand its functions and presumably build copies that perform the same movement patterns and therefore also can jump. Maybe we could build consciousness in a similar manner? By a careful procedure, we may perhaps map all the required details of the state of the robot to its ability to jump.

This procedure does not, however, catch the complete behaviour of the robot which should include the time-dynamics of its 'brain', being governed by the discrete logistic equation. Since this equation is chaotic, it is numerically impossible to work backwards indefinitely from a known, present state to accurately determine prior states that controls the robot's actions. For this reason a theory for the Jumping robot cannot be developed. Thus at best we can hope to build 'zombie robots' that copy the mechanical behaviour of the Jumping robots but not their associated 'brain' functions.

Moreover, our understanding of a *singular* emergent property or behaviour does not imply a complete understanding of *any* possible property or behaviour of the system. Having investigated all relevant aspects pertinent to the robot's jumping ability does not imply that we now can control the robot to, for example, jump on one leg only. *A posteriori* understanding of emergent systems is not sufficient for *a priori* understanding of their properties. The emergent states of the human brain, to be discussed shortly, must be understood similarly. A complete mapping of an individual thought process to its supervenient neural network certainly does not guarantee understanding of qualitatively different thought processes.

Furthermore, it is well known that mental processes suffer from an additional complexity, not related to emergence: they cannot be scientifically related to externally measurable properties in the same way as for the robots; the conscious properties of the mind are predominantly accessible internally or subjectively. Consciousness thus cannot be straightforwardly understood from basic externally controllable, non-emergent physical states, because the resulting characteristic emergent properties cannot be ascertained objectively. Our focus is here on emergence, so we will not dwell on this difficulty any further.

Let us now investigate the degree of emergence associated with consciousness. The human brain is vastly more complex than the discrete logistic equation. In brief, its approximately 100 billion neurons communicate as follows. Via the so-called dendrites, each neuron can be reached by electrochemical signals from tens to tens of thousands (average 7000) neighbouring neurons. The contributions from these signals are weighted in the neuron's cell body to an electrical potential. When this reaches a certain threshold, the neuron sends out a pulse along a nerve fibre, which in turn may connect via synapses to several other cell dendrites. The outgoing signal from a neuron is thus a step function rather than a continuous, nonlinear function of the incoming signal. This justifies the choice of the discrete logistic equation rather than its continuous counterpart in the thought experiment. Neurons fire about 100 signals per second with signal speeds of up to 100 meters per second. The behaviour varies from neuron to neuron. For networks of neurons, it is found that nonlinear functions called sigmoids, with S-shaped dependence of the input signals, provide realistic models. Mathematically, sigmoids are closely related to the discrete logistic equation.

We can thus conclude that communication within the neural network of the brain occurs nonlinearly and discretely with a complexity far exceeding that of the simple logistic equation. The behaviour of the latter is, as we have seen, characterised as algoritmically incompressible. Since, in a physicalist perspective, it is the neural processes of the brain that underlies consciousness, it follows that a theory for phenomenal consciousness cannot be constructed. Thus consciousness, and similarly subconsciousness, is *epistemically strongly emergent*. In the same way that we could not reductionistically explain the Jumping robot's behaviour, we can not reductionistically explain the properties of consciousness; the mind-body problem is unsolvable.

We may now ask whether consciousness is also *ontologically* emergent; are the properties of consciousness irreducible to the lower level states and processes that form the basis of consciousness, the ones that consciousness supervene on? The meaning of 'cannot be reduced to' for this question need be illuminated. Let us return to the example of the Jumping robot. The property to be able to jump is not ontologically emergent for the reason that in an objective meaning this property was an option within reach for the system, although its details were unknown to us. By 'objective' we mean that the various possible sequences of numbers being generated by the discrete logistic equation, of which at least one potentially lead to jumping behaviour, correspond to an amount of information that is manageable in principle. This latter statement demands clarification, since we now have made contact with the consequences of quantum mechanics and information theory for ontology.

It has been shown (Lloyd 2002, Davies, 2004) that the information storage capacity of the universe is limited by the available quantum states of matter inside the causal horizon. The latter is the distance, limited by the finite speed of light, outside which no events may be causally influential. Using multiple arguments it is found that the order of 10^{120} bits of digital information may be contained within this horizon. The fact that this result is an estimate is not essential; what matters here is that information storage capacity is limited to a magnitude of this nature. For a property or behaviour to be characterized as ontologically emergent it must therefore exhibit a complexity transcending some 10^{120} bits of digital information; only then can we say that there is no possibility, even in principle, to 'reduce' it to the low-level phenomena on which it is based. Note that ’ontology’ is used in this work in the traditional, philosophical sense and *not* as a reference to properties or interrelationships between entities used in computer science and information science.

It would certainly be helpful if we could find a specific example of an ontologically emergent system. To this end, we note that emergence rarely is associated with the results of human activities, with *design*, but rather with *evolution*; the development of nature. Evolution has through natural selection access to a tremendous diversity of degrees of freedom and features a huge potential to generate emergent systems. An example from chemistry is myoglobin, an important oxygen binding molecule found in muscle tissue (Luisi, 2002). Here 153 amino acids are interconnected in a so-called polypeptide chain. Since there are 20 different amino acids, the number of possible combinations of chains amounts to the enormous number $20^{153} \approx 10^{199}$, which corresponds to a number of digital bits much larger than 10^{120}. Myoglobin is thus an ontologically emergent phenomenon in nature; the molecule is, in terms of optimized functions such as high oxygen affinity, not reducible to its low-level constituents. It could only evolve, it could not be designed.

Returning to consciousness, we will now argue that individual thought processes are emergent. We recall that the example of the Jumping robot was used for illustrating strong epistemic emergence; a certain robot property (or behaviour or state) X was to be designed but our modelling capacity, even using computer simulations, turned out to be insufficient to accomplish X. This is precisely what we mean by strong epistemic emergence; high-level behaviour could not be predicted or controlled from low-level behaviour. However, since the possible number of combinations of robot joints, muscles and other bodyparts is limited we have reason to assume that the computational capacity of the universe, if accessible, would suffice for solving the problem. Thus X is an example of strong epistemic but not ontological emergence. Next, we found an instance Y of ontological emergence in the myoglobin protein molecule. Its high oxygen affinity is ontologically emergent because of its extreme complexity. Just as X surprisingly appeared among the robots on the island, Y develops as a ’surprise’ to nature itself.

Let us proceed along the same lines as earlier: we wish to determine whether consciousness entails any property Z that is ontologically emergent. More precisely: given the physical resources on which consciousness supervenes, is it possible to reduce all occuring properties or behaviour of the mind ontologically to these low-level processes? Is the information processing capacity of the universe sufficient for this task? Let us imagine a human-like individual in a situation where conscious choices need be made. This could, for example, be to select courses from a menu at a restaurant. The final choice will be influenced by factors such as earlier taste experiences and attained preferences, allergies, views on appropriate diets, present state of hunger, the internal matches between starters, main courses and desserts, memories from possible earlier visits to the restaurant, views on suitable accompanying drinks as well as prices. Just as we interested us for the jumping ability of the robots we are now interested in the individual’s choice of a three-course dinner. What are the required low-level conditions for that the individual should have the property of choosing a certain, specified combination of these dishes? We will now argue that this property is an example of Z.

Assume that one-thousandth of the 10^{11} neurons of the brain are involved in the associated thought process. Assume also that there is some redundancy, or cluster behaviour, so that only one-thousandth of these in turn need to actively be controlled. This leaves $N = 10^5$ neurons to be initiated. It can be assumed that their states are binary; only two states exist for each. Thus there are 2^N possible combinations of neural states. Clearly even a small fraction of these possible states is much larger than 10^{120}. Similarly as for the oxygen affinity of the myoglobin molecule it is not possible to *design*, even in principle, a thought process that would correspond to, for example, the choice of a particular three-course meal. This extremely complex mental process has evolved to take place at a level that is not reducible to low-level processes. As a result thought processes of the mind generally are ontologically emergent; they cannot be reduced to the low-level components, the neurons, that they supervene on. As we will argue shortly, this opens up for 'downward causation'.

Needless to say, we could imagine even more complex examples of properties Z, including those related to subjective experiences.

3.4 Neuroscience

The neural networks of the brain communicate in discrete nonlinear processes to generate cognitive functions such as the abilities to feel pain, think, make choices and introspect. If these basic processes were linear, their behaviour could possibly be reduced to a theory. This theory would have lower complexity than what it describes since it would be algorithmically compressible. Nonlinear systems like the neural network of the brain, however, generally feature higher complexity. Since the neural network associated with consciousness is nonlinear to its nature and epistemically strongly emergent, we have argued that a theory cannot be constructed for it. Thus consciousness cannot be understood within a reductionistic framework. This holds regardless if we seek a computational theory of mind or any other formally reductionistic theory of mind.

It is of interest here to briefly discuss a quite different obstacle for understanding consciousness. Abandoning efforts for finding theories of consciousness, we may be inclined to instead turn to the possibility of artificially designing consciousness. In neuroscience there is a search for 'neural correlates of consciousness' (NCC), which form the neural processes in the brain that are directly linked to the individual's current mind activities.

Let us say that NCC:s indeed can be identified to an extent that serious attempts to create conscious processes in artificial brains can be made. On each such experimental attempt the function must be ensured; the system must be diagnosed. Otherwise there is the possibility that we have designed an advanced system that externally behaves like a consciousness but actually lacks mental processes. The problem is here that essentially no fundamental limit exists for non-cognitive 'intelligence' of advanced computer programs. These would then, properly designed, be able to pass any kind of Turing test. In these tests, where the respondent is hidden so that the person performing the test does not know whether it communicates with a human or a machine, any machine that produces similar responses as humans when addressed are deemed intelligent on the level of a human.

The Turing test is valuable for testing intelligence, but obviously is unreliable for testing consciousness. But what would then be an adequate diagnosis? Current definitions of phenomenal consciousness provide an answer: we are to ensure that the system can have subjective experiences. But since all measurement of the functions of consciousness must be done externally, that is by laboratory personnel using diagnostic equipment, the system's internal cognitive functions cannot be measured. There is simply no information externally available from the system that would be indistinguishable from that which can be produced by an advanced, but unconscious, computer program. We could be facing an intelligent robot, without ability for conscious behaviour. This is, as mentioned elsewhere, therefore not a viable route for solution of the mind-body problem.

In short: understanding of a system implies the possibility of constructing it, with all of its functions. But since the intended function (generation of conscious thoughts) of the systems we may construct cannot not be experimentally verified, we can not say with certainty that we understand them.

4 Consequences for free will

Must we have the thoughts we have? Do our thoughts only happen, rather than being created by ourselves? Does determinism hold our will into an iron grip? The free will problem presumably is the most important existential problem and has generated shelf kilometers of literature throughout the centuries. One reason for the problematic situation is very likely the most common definition of free will itself: 'the ability to act differently'. Indeed, it is hard to see any opportunity for scientific methods to determine whether we actually can 'act differently' or not. How do we know whether an individual's actions are autonomous or predetermined? And why should even a free consciousness act differently in two identical situations? Many arguments about the will thus lead to uncertain terrains.

4.1 Alternative definition of free will

We here make an attempt to provide a definition of free will with characteristics that are scientifically decidable: *A conscious individual has free will if its behaviour takes place according to its intentions, the intentions are not subconsciously generated and if the individual's mind is an ontologically open system.*

By 'will' we here refer to expressions by a cognitive system of preferenced future actions. Furthermore, by an 'ontologically open system' is meant a causal, physically closed system for which the future behaviour is undeterminable and thus uncontrollable, even in principle. We will justify the definition as follows. Experience tells us that basic, lower level phenomena are causal and deterministic. Quantum mechanical statistical corrections must, of course, be taken into account as discussed below. If also the higher level neuronal functions of the conscious individual are determinable beforehand, it is obvious that its will is governed by laws outside its conscious control. This circumstance is a feature of the classical, deterministic argument against free will. Behaviour related to ontologically open conscious systems is, on the other hand, not directly reducible to earlier physical neural states. As argued in the following, this is typical for ontologically emergent behaviour. Individuals featuring ontologically emergent neural processes are ontologically undeterminable beforehand. They cannot be externally controlled, not even in principle. This is the crucial and decisive property we assign to ontologically open systems. Noteworthy is that ontological openness is possible also for systems that are deterministic (or essentially so) at lower-level, as shown below. Consequently the notions of compatibilism and incompatibilism are irrelevant here.

Phenomena relating to classical *open physical systems* are generally causal, but indeterminable. These systems are open to *external influence*, and they are thus not guaranteed to evolve identically when repeatedly started from the same initial conditions. The associated dynamic processes should *not* be regarded as random or chancy; the point is that the system itself does not contain sufficient information about its future external states. This becomes clear if we extend the size of the system to also include all of its external influences. Such a system may be physically closed, causal and deterministic. We will, in the next section, argue that consciousness is ontologically open in spite of being a physically *closed* system.

For the sake of completeness and accuracy we should, as mentioned, account for the implication of quantum mechanics that determinism does not fully apply at the very micro-level; the uncertainty principle shows that nature is 'blurry' at the sub-atomic and atomic particle levels. For larger clusters of particles, however, like the molecules that make up the neurons, this effect is of much less importance, because of so-called quantum decoherence. The concept of 'adequate determinism' has been coined to emphasize that the statistical determinism that results and is used here, in essence is correct in the macroscopic world, even if quantum phenomena are important for very small systems.

Returning to the present definition of free will, we note that the desired actions of a free consciousness must not turn into anything other than intended; *behaviour must be based on its intentions*. If I wish to consider what to eat for dinner, such a reflection must be possible. My actions must consistently and adequately follow my will. The phrasing 'takes place according to its intentions' is deliberately somewhat vague; the precision we may strive for in our actions is sometimes not achieved; this is not because the will is not obeyed but rather from our physical and psychological limitations. Note that we also assume conscious individuals; it is not meaningful to talk about 'will' for other systems.

Finally, the condition that '*the intentions are not subconsciously generated*' is needed to ensure that the individual's 'brain' does not contain any hidden systems that manipulates it in a manner that consciousness, in spite of being controlled, experiences intentions as its own. So-called 'character decisions', decisions that we make without active reflection based on our experiences and consolidated positions, we treat in this context as conscious. We will return to these.

In summary, we have cast the formulation 'the ability to act differently' into an alternative, scientifically decidable formulation in order to generate improved methodological conditions to handle the free-will problem. A condensation of the definition could be something like 'An individual that can realize conscious, unforced choices has free will'.

The task is now to address the inhibiting circumstance that the mind must feature a deterministic character in order to enable coherent thought processes and consistent performance of its intended actions, but at the same time feature an ontologically open future to permit self-caused actions. Free will implies that this potential contradiction is dissolved. It is at this point the emergent character of consciousness becomes crucial.

It should be noted that if it really were the case that a true theory of consciousness could be designed, then there is no room for free will. *Free will implies that the mind is epistemically strongly emergent*, a fact that deserves more attention in the literature. The reason is simply that if the individual's behaviour would be epistemically computable or could be simulated, its behaviour would be predictable and controllable and thus not free.

4.2 Consciousness and determinism

We will now argue that consciousness is an ontologically open, high-level global system. The question is then how it would be possible for the brain's essentially deterministic atomic lower-level activity to lead to ontologically open behaviour at the higher inter-neural levels related to consciousness, considering that man and consciousness are of the physical world.

At this point it can be instructive to discuss a hypothetical Jumping robot. It is clear that if its functions were not epistemically emergent, we could control it. We could then device certain input to its 'brain' in order to produce a certain behaviour, like jumping. If the robot furthermore had conscious capabilities we could in principle fool it into believing that it chooses to jump out of free will by manipulating its memory and cognital functions. If its functions instead were characterised by epistemically strongly emergent behaviour, as in our thought experiment, it cannot be controlled. Consequently this robot is, in a sense, free in relation to us. Moving to an even higher level of complexity, we may next assume that the robot's behaviour is truly ontologically emergent. This extends the freedom of the robot since its future actions are undeterminable, even in principle. It may be objected, however, that there is no escape from the robot's deterministic dependence on its initial set-up and conditions in combination with the laws of nature. If the robot is repeatedly started from exactly the same conditions, its behaviour would be the same. This contradicts our intuitive perception of free will. We would certainly hesitate to assign free will to this robot.

Experience shows that causality applies in our physical world. This means that the current state of a typical physical system, in terms of the positions and velocities of its microscopic constituents, provides a sufficient condition to take it to a subsequent state; cause results in effect. We endeavour to express the regularities of cause and effect as laws of nature. If new physical states can be found uniquely from previous states of the system, we talk about determinism. Stated equivalently: determinism implies that the evolution of a system, if repeatedly started from the same initial conditions, will always be identically the same. Everyday events, such as when the billiard cue hits the cue ball which subsequently knocks down the yellow ball in the hole, tempt us to believe that causality and determinism are equivalent concepts. But they are not. The future of a specified causal physical system may actually be undetermined, even disregarding the statistical nature of quantum mechanics. This happens when the system is open in some sense, that is when unaccounted external phenomena may have an influence.

Our everyday experiences could lead us to the erroneous conclusion that change and the underlying forces act instantaneously. All effects of forces in nature, that is changes of state, are in fact due to combinations of the four basic forms of interaction through exchange of particles called bosons. This interaction takes indeed a finite time; the lower limit is given by a key relationship in quantum mechanics called Heisenberg's uncertainty principle (Lloyd, 2002). The limiting time is proportional to Planck's constant and inversely proportional to the system's average energy above the ground state, which for a one kilo system means that no more than $5 \cdot 10^{50}$ changes of state are possible per second.

In the entire visible universe, which dates back some 14 billion years and has a mass of 10^{53} kg, in total a maximum of about 10^{121} changes of quantum states have occurred. Although this is a huge number it is not infinite. The universe's 'capacity to act and compute' is thus limited by Planck's constant and the available energy. As a consequence, the futures of very complex physical systems are 'unknown' even to nature at each instant in the sense that this information cannot be processed and obtained presently; it requires too much computational resources (Wolpert 2008). The fact that future states of very complex systems are ontologically undecided is of course important when discussing determinism and the relation to ontologically open systems.

Let us now consider the behaviour of a hypothetical single conscious individual placed in a closed room, without contact with the outside world. Also assume that there is, in an objective sense, complete access to all information relevant to the individual including all initial and boundary conditions. We are interested in the character of the individual's behaviour in a certain future time interval. For the sake of argument let us first consider an imagined case of behaviour that we could deem as undeterminable with respect to the individual's choices and actions. If the individual, before taking a decision, had the magic ability to consult a clever genie inhabiting some dimension otherwise unrelated to our physical world, the individual's future would clearly not be deterministically given. The influence of the genie's advice on the individual's behaviour would be similar to the case of external signals influencing an open physical system. Since the individual's decisions are not immediate consequences of its present physical state, we must infer that the will of this individual is not limited by any deterministic dependence on its initial set-up and conditions. In discussions of determinism, in a similar vein as that of Laplace in *Essai philosophique sur les probabilities* (1814), it is often asserted that given the positions and velocities of all particles in the universe, its future would be in principle determinable. It is sometimes forgotten, however, that the argument implicitly assumes the continual action of given laws of nature. The appearance of the genie violates this assumption.

Returning to reality the genie of the thought experiment can, with the same result, be replaced by the individual's emergent thought processes in combination with preferences being acquired during its earlier life, now stored in its memory. Will is about planning and experiences are necessary prerequisites. These experiences are personal and rated subjectively, whereafter they are remembered and used as a basis for subsequent preferences. The preferences are consulted, similarly as the genie, before decisions are taken. Furthermore the preferences are the result of ontologically emergent processes where subjective positive or negative connotations have been related to various events, actions and choices.

Alternatively formulated, consciousness, in any given time interval, acts as an open system in the sense that the memories associated with subjective preferences are ontologically unrelated to the physical situation. The fact that in principle it is possible to, atom by atom in a Laplacian sense, build the individual's entire network of coupled neurons is not relevant here. Such a system would still have built in subjective preferences, the character of which would be unknown and, as we have argued, with functions equivalent to conferring with an independent genie. Ontological emergence is again crucial in that it decouples the physical state of the system from its subjective properties and behaviour. This concludes our argument for that behaviour of conscious individuals is ontologically open.

Summarising, consciousness is an ontologically open system and thus undeterminable and uncontrollable in principle. We have argued that it is meaningless to say that the evolution of consciousness is deterministically given since at any instant ontologically emergent subjective experiences, stored as memories in the mind, govern the individual's behaviour.

How could, we may wonder, such complex behaviour evolve in humans? Perhaps the most competitive evolutionary aspect of consciousness is its ability for planning in order to avoid dangers, gain advantages and to optimize longer time survival. Planning requires alternatives to compare with. The alternatives manifest themselves to us humans through experience; we are by necessity not born with fixed perceptions about the world since, for example, our environments differ depending on where we are born. Our experiences need storage, or memory, so that they can become conscious alternatives when we are about to make choices. Together with these objective experiences, we have also stored associated subjective impressions. In the process of planning, when making our choices, it is precisely the subjective impressions that influence our decisions. These are so-called 'character decisions', discussed in the next section. Again, ontologically emergent and thus externally uncontrollable brain processes act so as to store personal and subjective impressions for subsequent use in decision making processes. This is certainly in line with our notion of free will.

The effect of memory on consciousness to continuously modify it has the result that consciousness may act differently, even if external conditions are unchanged. Repeatedly facing identical external situations, conscious individuals can make new and different choices each time, as a result of recollections of subjective experiences of earlier instances.

We have, from a physicalistic and thus monistic position, argued for that the mind is an ontologically open system. Interestingly, the same result seems to follow from a dualist perspective. To show this, assume for a moment that dualism holds; there is both a material and a somehow separated ’mental dimension’. What characterises activity in the mental dimension? Certainly not randomness; scientific analysis of mental behaviour speaks against this. But if the mental dimension features regularity and law-bound processes we face a similar question as when taking the physicalistic stance: what is the maximum freedom that can be excerted by the will, given the laws of nature? Thus a natural conclusion is that dualism cannot provide conscious will with higher degrees of freedom than those found within physicalism.

To conclude, the mind's ontologically open high-level character is argued to be a result of the ontologically emergent nature of the causal processes in its complex neural network, combined with the limitations in nature's capacity for deterministically connecting present states to future states.

4.3 Willed intentions and the role of subconsciousness

Free will requires, according to the present definition, that individual behaviour takes place *according to the individual's intentions*. This condition is not really problematic; it is satisfied by our experiences. The individual's everyday functioning is completely dependent on that she consistently carries out what she decides. Does she want to make herself a cup of coffee, she does it. The exceptions that can be identified, such as shortage of coffee or that she cannot be bothered, are not about principal mental limitations but of physical or psychological disability or of limitations in the outside world.

So far, we have presented arguments for that the combined consciousness/subconsciousness meets the requirements for free will. But few would regard this as sufficient; if our volitional decisions, in spite of their onthologically open origin, are unconsciously dictated to us it is difficult to speak of free will. There is evidence, however, that consciousness in a number of situations exerts its will without significant influence from the subconscious. First, it should be noted that there is a spectrum of degrees of collaboration between the two. Our experiences of dreams show that subconsciousness may be active when we are not consciously aware. Driving a car along a well-known road is a well known example of collaboration between consciousness and subconsciousness. And participation in an intense discussion is an example of consciousness mainly acting on its own. But the independent role of consciousness and the will has been strongly questioned over the past few decades and some authors talk of "the illusion of free will". Support is partly found from neuroscience. A "readiness potential," being activated unconsciously well before we make conscious decisions, appears to reveal that the main decision-making takes place beyond consciousness. A pioneer in the field was Libet (1985). The experiments since then carried out within this field has, however, many possible sources of error and have endured severe criticism in several places (Klemm, 2010 and Baumeister et al, 2011).

In certain practical situations it is, from an evolutionary point of view, crucial that consciousness may act undisturbed. The need for rapid and well balanced decisions, as when we are driving and we suddenly need to consider how to avoid a car that suddenly wobbles into the roadway, is one example. In a very short time we need to perform a large number of considerations, including how to avoid colliding with people while at the same time ensure our own safety. The subconscious mind would not, with the associated delay that Libet's and other experiments show, have time to gather all the relevant information in order to survey the situation and in a short time deliver adequate decisions that do not conflict with our conscious perception and handling of the situation. Certainly, if conscious decisions would not be important in situations like these, evolution would have provided us with a mechanism that automatically disconnected consciousness in favour of subconsciousness, in a similar manner as when we react reflexively.

Furthermore it is well known that, upon learning new knowledge and skills, performance is gradually taken over by the subconscious as we become more knowledgeable and skilful. But for the beginner who sits down at a piano, the subconscious mind is completely unprepared. There is no way for the subconscious to control the finger movements because it does not 'know' what should be done (Klemm, 2010). Obviously more research is needed to identify to which degree subconsciousness impacts on our actions, however we have here given examples where the subconscious cannot reasonably have a significant role.

The cooperation between consciousness and the unconscious points to a second argument why consciousness must not be controlled by the subconscious. Neuroscience shows that a significant part of the 'processors' of the brain used for conscious thought are also used for unconscious processes (Dehaene, 2014). This provides support for that not only global neural processes of consciousness but also those of subconsciousness are ontologically emergent. Thus, whereas deterministic processes (in the adequate sense, as discussed above) contribute to communication between consciousness and the unconscious, the latter is not deterministically controlling consciousness. As pointed out, experience shows that we can consciously cancel impulsive intentions, using our "free won't" (Libet, 1985).

From another perspective, we do not necessarily need to make a distinction between consciousness and the subconscious as clearly separated global systems. Already individual neurological subsystems associated with the mind appear to be sufficiently complex to render their interaction ontologically emergent. In the subject of game theory similar results have, interestingly enough, been found. Emergent behaviour has been observed in simulations of nonlinear interaction between two players, who both act in order to optimize their game while trying to act unpredictable for the opponent, if players are allowed to make use of the game's history (West and Lebiere, 2001).

A complication related to the distinction between subconscious and conscious choices is what we might call 'character decisions'. Based on previous experience and reflections, people accumulate different positions or traits of character that could lead to routine behaviour in certain situations. Facing an approaching threatening individual, for example, certain people will normally escape while others preferably stay to deal with the danger. This behaviour does not necessarily constitute an active conscious choice of the type we have discussed so far, but may rather be a result of the individual's disposition to act in such situations. Since the individual normally is aware of her traits of character, we here consider the nature of character decisions to be conscious rather than unconscious.

Our feelings, thoughts and choices do not simply happen to us. They arise from basic neural processes related to our minds and are developed emergently in a cooperation between consciousness and the unconscious. But how, then, can our thoughts and feelings take form in a structured and coherent way? How can the individual carry out her intentions unruled by the subconscious? These important questions are not analyzed here; of prime interest for the question of free will is that thoughts and feelings arise in a manner which is neither determinable nor controllable.

5 Discussion

The results of this work unequivocally point towards non-reductive physicalism; mental states supervene on physical states but cannot be reduced to them. We may ask what the consequences are for causal closedness, that is the thesis that no physical events have causes beyond the physical world. Our answer is that the physical world really is causally closed in the ontological sense because we have every reason to regard it as causal; every physical state leads, in accord with the laws of nature, to a new state. For simpler, low-level systems, new states are in principle predetermined and sometimes even computable. The human brain employs deterministic low-level processes at the neural level for thought processes, carrying out certain actions (somatic nervous system) and for reflexes (autonomic nervous system). But as we have shown, this does not mean that all systems in the physical world are deterministically controlled. Emergence can alter the situation. Consciousness, which we have shown to be ontologically emergent, is such a high-level system. From an epistemic perspective this means that the possibility for conclusions about the causal functionality of mental systems are limited. The situation is reminiscent of that of mathematics for which Gödel proved that there are true theorems in the system that are unprovable because of its complexity.

We can now explain why emergence does not cause *overdetermination* with regards to the causal situation for consciousness (Kim, 2006). It has been argued that if the dynamics of consciousness is determined by its current state plus the laws of nature, then emergent phenomena cannot exist independently; they must be a result of the conditions already provided. Otherwise we seem to be facing an overdetermined problem. The solution to the dilemma is that the emergent properties are of the same nature as those new conditions that may present themselves when a closed system is transformed to an open system. They are thus additional conditions, being governed by associated additional relations. Mathematically speaking, just as many new equations are added as new variabless. Thus overdetermination is avoided. For the example of the person being in a closed room, this could correspond to the door being opened. Emergent properties have thus, as far as deterministic control is concerned, the same impact on the development of the system as external influences have on an open system. This solution to the problem of overdetermination also explains how 'downward causation' (Kim, 2006) can take place. Interacting emergent phenomena can determine the development of the system (in this case, the mind) independently of the causal situation at lower levels.

What is then the implication for compatibilism; the position that determinism is compatible with free will? Interestingly, whether compatibilism or incompatibilism holds is not relevant here. Even though neural low-level processes are deterministic, high-level cognitive phenomena are ontologically indeterminable and uncontrollable in spite of being defined by low-level processes. The point of transition is, as we have seen, governed by the amount of information that can principally be stored and processed in a physical system. This also illustrates the point that in a physicalistic view of the world, determinism or indeterminism has no bearing its reductive character; it is emergence that renders physicalism nonreductive.

Finally, how do these results relate to epiphenomenalism, the notion that mental states are only by-products of the physical states and unable to causally influence these? To answer this question, we need to note that the form of non-reductive physicalism assumed here is not a form of property dualism. Although mental states are not deducible from basic neurological states, they certainly directly correspond to physical states; they supervene on these. Non-reductionism follows because of the emergent character of mental states, not because of lack of correspondence between physical and mental states. Thus epiphenomenalism is ruled out here.

6 Conclusion

In a physicalist analysis of the mind-body problem, resting on results from mathematics and physics, the concepts of algorithmic information theory and emergence are used to argue that the problem is unsolvable. The vast neural complexity of the brain is the basic obstacle; from a thought experiment it is shown that even a much simpler but related nonlinear system may exhibit epistemically strongly emergent properties. Reductionistic understanding of consciousness is thus not possible. Neuroscience will continue to make progress - we will almost certainly find, for example, the cognitive centra that are active at certain stimuli or thought processes - but emergent cognitive phenomena like qualia or introspection will not be expressed in a theory. The 'explanatory gap' cannot be bridged.

From a scientifically and methodologically more useful definition of free will than the traditional it is subsequently argued that high-level neural cognitive processes are ontologically open, even though underlying physical laws and lower-level processes are deterministic. As a consequence conscious processes are not determinable, not even in principle. Thus, the three requirements for free will suggested here are satisfied; that the individual's actions take place on the basis of its intentions, that these intentions have not been subconsciously forced onto the individual and that the individual behaves as an ontologically open system.

Acknowledgements

Many thanks go to Mr Keith Elkin, for sharing his knowledge in neuroscience and several related concepts as well as for comments during many philosophical discussions. Thanks also professor Erik J. Olsson for insightful and constructive discussions on several aspects of the work.

References

Baumeister, R. F., Masicampo, E. J., & Vohs, K.D. (2011). Do Conscious Thoughts Cause Behavior? Annu. Rev. Psychol., 62, 331-361

Block, N. (1995). On a confusion about the function of consciousness. Behavioral and Brain Sciences, 18, 227-247

Carnap, R. (1950). Logical Foundations of Probability. (University of Chicago Press)

Chaitin, G. J. (1987). Algorithmic Information Theory. (Cambridge University Press)

Chalmers, D. J. (1995). Facing up to the problem of consciousness. Journal of Consciousness Studies, 2, 200–219

Chalmers, D. J. (2006). Strong and weak emergence. The Re-Emergence of Emergence. (Oxford University Press)

Chalmers, D. J. (2016). MindPapers. http://consc.net/mindpapers/

Davies, P. C. W. (2004). Emergent biological principles and the computational properties of the universe. arXiv preprint astro-ph/0408014, 2004 - arxiv.org

Dehaene, S. (2014). Consciousness and the Brain. (Penguin Books, New York)

Kim, J. (1999). Making sense of emergence. Philosophical Studies, 95, 3-36

Kim, J. (2006). Emergence: Core ideas and issues. Synthese, 151, 47-559

Klemm, W. R. (2010). Free will debates: Simple experiments are not so simple. Advances in Cognitive Psychology, 6, 47-65

Levine, J. (1983). Materialism and qualia: the explanatory gap. Pacific Philosophical Quarterly, 64, 354-361

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. Behavioral and Brain Sciences, 8, 529–566

Lloyd, S. (2002). Computational Capacity of the Universe. Physical Review Letters, 88, 237901-1-4

Locke, J. (1690). An Essay Concerning Human Understanding

Luisi, P. L. (2002). Emergence in chemistry: chemistry as the embodiment of emergence. *Foundations of Chemistry*, 4, 183-200

McGinn, C. (1989). Can we solve the mind-body problem? *Mind*, 98, 349-366

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review,* 83, 435-450

Silberstein, M. & McGeever, J. (1999). The search for ontological emergence. *The Philosophical Quarterly*, 49, 182-200

Van Gulick, Robert, "Consciousness", *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2014/entries/consciousness.

West, R. L. & Lebiere, C. (2001). Simple games as dynamic, coupled systems: randomness and other emergent properties. *Journal of Cognitive Systems Research*, 1, 221-239

Wolpert, D. H. (2008). Physical limits of inference. *Physica D* 237, 1257–1281