Free Will of an Ontologically Open Mind

Jan Scheffel
Division of Fusion Plasma Physics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden, jans@kth.se

Abstract

The problem of free will has persistently resisted a solution throughout centuries. There is reason to believe that new elements need to be introduced into the analysis in order to make progress. In the present physicalist approach, these elements are emergence and information theory in relation to universal limits set by quantum physics. Furthermore, the common, but vague, characterization of having free will as "being able to act differently" is, in the spirit of Carnap, rephrased into an explicatum more suitable for formal analysis. It is argued that the mind is an 'ontologically open' system; a causal high-level system, the future of which cannot be reduced to the states of its associated low-level neural systems, not even if it is rendered physically closed. We believe that a positive answer to the question of free will is outlined. Downward causation is supported and Kim's problem of causal overdetermination is resolved.

Keywords

Free will, determinism, downward causation, emergence, ontologically open, mind-body problem, consciousness, subconsciousness.

1 Introduction and background

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 can be traced to the common definition of free will as '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. For example, why should even a freely acting consciousness behave differently in two identical situations?

In Scheffel (2019) it is argued that consciousness cannot be represented by a theory and, as a consequence, that the mind-body problem is unsolvable. The associated *epistemological emergence* of consciousness is of interest for the problem of free will since if, on the other hand, a theory for consciousness could be designed, then its behaviour would be computable or could be simulated. It would thus be in principle predictable and not free; if we could understand consciousness there would be no room for free will. It is somewhat surprising that this relationship is not much discussed in the literature. The argument was subsequently carried a step further to show that consciousness, as a high-level property of the mind, is *ontologically emergent* with respect to its low-level neural states. The following characterization was used: a high-level property is ontologically emergent with respect to properties on low-level if the latter form the basis for the high-level property and if it is not reducible to properties at low-level. Following van Riel and van Gulick (2018) *ontological reduction*, in turn, should entail "identification of a specific sort of intrinsic similarity between non-representational objects, such as properties or events". An ontologically irreducible property, if it exists, hence could not be determined by its low-level-properties or behaviour; it could not be characterised by a statistical or law-like behaviour in relation to its low-level components. In a sense its behaviour comes as a surprise to nature.

It was further found that extremely complex systems may feature ontologically emergent properties. The reasoning is based on elements of algorithmic information theory (Chaitin 1987) and the ontological quantum mechanical limits for information and computational capacity (Lloyd 2002 and Davies 2004). An elucidating example is the molecule myoglobin. It consists of a polypeptide chain of 153 amino acids. On each position of the chain, there could be found either of 20 amino acids, rendering as many as $20^{153} \approx 10^{199}$ possible combinations. Since this number by far exceeds 10^{120}, the approximate number of possible quantum mechanical states of the universe, the high oxygen affinity of the molecule could not be computed, even in principle. As a property it could not be designed, it could only evolve.

The argument proceeds by showing that although low-level neural states form the basis for consciousness, consciousness is not ontologically reducible to the properties of these because of the extreme complexity of the cortical neural network. A main result is that if properties of a complex system, being the result of for example long term evolution, can only be manifested by the system itself - that is if nature for reasons of limited information storage capacity cannot accommodate a formal representation of the system's properties - then the system features ontologically emergent properties. In this paper we argue that the degree of freedom resulting from the ontologically emergent character of consciousness dissolves the deterministic difficulty we have been facing for freedom of the will.

An argument for free will must consider causal closure and physical determinism (Popper and Eccles, 1977). Assuming causality, causal closure is the position that no physical event, like a decision formed in our brain, has a cause outside the physical world. Physical determinism, or simply determinism, says that a system's future is fully determined, or specified, by its present state. We will touch upon microscopical uncertainties caused by quantum mechanical effects later.

The concepts of causality and determinism can be interpreted by considering the order of related events in time. Causality is *a posteriori* in the sense that it implies that any event of a physical system can be traced backwards in time, as the result of one or more causes. In the sciences, this enables interpretation. If the system is physically closed, so that it does not interact with the external physical world, all potential causes for future events are contained within the system itself. The future evolution of such a system may be implied *a priori*, in which case it is deterministic. In the physical sciences, this in principle enables prediction through the use of theories, like natural laws or simulations. Determinism is usually implicitly assumed in the physical sciences. An open physical system may, by definition, interact with the world external to the system. While preserving causality, it cannot be assumed deterministic. Causality does not imply determinism since causality does not require particular, individual causes to uniquely specify the future of the system.

In this physicalist approach it is assumed that causal closure holds. We will also assume that if an isolated physical system with cognitive properties somehow could be started repeatedly from *identical* initial conditions, so that also its internal memory of its history is the same, it will evolve identically (within quantum mechanical uncertainty limits) in all cases. However, we will argue that this complex system is *not* deterministic in the traditional sense. The reason is that the time evolution of the system is not reducible to the properties of its low-level components alone. The causal status of the system is found from the simultaneous interaction between both low-level and emergent high-level phenomena, or properties. A cognitive system of this complexity may accommodate high-level agents, including subjective intentions, that lead to behaviour associated with free will. This is the core argument that will be developed in this paper, relieving the tension for free will that comes with the assumption that determinism at low-level implies low- to high level determinism. The argument will be outlined in section 3.

Are similar positions found in the previous vast literature on free will? Since we believe that emergence is a required element of a solution to the problem, the number of related articles is relatively limited; Stephan (2010) is an interesting exception. Even in some well known modern accounts of free will, the role of emergence is not identified; see for example, Dennett's and Wegner's influential works (Dennett 1997, Wegner 2002). In the Stanford Encyclopedia of Philosophy (O'Connor and Franklin 2018) emergence in relation to free will is essentially neglected. The concept of emergence is, however, present in several discussions of consciousness and the mind-body problem (Kim 1999 and 2006, Chalmers 2006 to name a few).

Recently List (2014, 2019) has proposed a theory in support of free will. Whereas he avoids explicit reference to emergentism, the analysis is based on a separation between free will, as a ”higher-level” phenomenon found at the level of psychology, and fundamental physical ”lower-level” phenomena. In his ”Why free will is real” (2019), an extensive literature study has been carried out; the reference list also contains recent literature on the problem of free will. For the present work, additional references of interest are Campbell's (1974) introduction of the concept of 'downward causation', Kim's skepticism against emergence and downward causation (1999, 2006, 2011), recent defense of downward causation (Campbell and Bickhard 2011, among others) and arguments for causal efficacy without downward causation (Macdonald 2007).

As in the present work, List sharpens the characterization of free will and contends that high-level mental phenomena supervene on lower-order physical processes but are irreducible to this base. According to List, free will implies intentional, goal-directed agency, alternative possibilities among which we can choose, and causation of our actions by our mental states, especially by our intentions. For the latter requirement to hold, emergence of consciousness and will (”intentional action”) is required. The arguments supporting emergence and the effect of emergence in relation to free will have, however, been criticized as too weak (Weissman 2019, Bonilla 2019). It is, for example, not shown in any detail why mental states, as emergent, are irreducible to physical, neuronal states. The argument for how a system that behaves deterministically at a ”micro-level” can behave indeterministically at a ”macro-level”, and thus according to the author enable free will, is not fully convincing. Furthermore it is not clear how ”thinking and intending” as ”properties of the mind, not of the brain” can account for mental causation.

We here approach the role of emergence in relation to free will somewhat differently. We will assume that consciousness, as a property of the neural network of the cerebral cortex, is ontologically emergent. After having introduced a characterization of free will, intended to lend itself for analysis in a somewhat stronger sense than more traditional characterizations, we will need to ascertain whether consciousness, as an emergent system, belongs to the class of 'ontologically open systems'. The positive outcome of this analysis helps to overcome the potential straitjacket with respect to alternative possibilities for intentional action, due to supervenience of conscious high-level processes on deterministic processes at low-level, being problematic for List (Bonilla 2019).

We start in section 2 by discussing a common characterization of free will and introduce a slightly modified definition. In order to explain how ontological emergence renders determinism consistent with free will, the concept of 'ontologically open' systems is introduced and discussed. These systems are causal high-level systems, the future of which cannot, even in an *a posteriori* sense, be reduced to the states of their associated low-level-systems, not even if these systems are physically closed. In section 3 we argue that consciousness, being an ontologically emergent system, is indeed also ontologically open. Thus it is found that consciousness satisfies all three conditions required for free will according to our definition. In section 4 we discuss the role of subconsciousness. The paper ends with a discussion and conclusions.

2 Alternative definition of free will

Common characterizations of free will like, for example, 'ability to act differently' are problematic. How would we resolve the question whether consciousness has an 'ability to act differently'? What information should be found? Consider, for a moment, the question "Was Hamlet left-handed?". This is a proper, well-defined question, but we clearly see a problem. If Shakespeare never discussed Hamlet's handedness, nor mentioned or hinted at it in any of his works, there is no information available to resolve the question, ever. There is a 'gap' between the information asked for and accessible reality. In 'the ability to act differently', 'differently' is about outcomes, which can be identified experimentally. But 'differently' also refers to the neural processes that deliver the outcome of the agent's deliberation. These could be of strictly deterministic, low-level origin or be associated with emergent high-level, conscious considerations. As we will see, the degree of freedom for the will is quite different for these two cases. Moreover, 'ability to act' concerns a cognitive and subjective first person process to which we have no third person access, neither theoretically nor experimentally. There appears indeed to be an unbridgeable gap that cannot be crossed in order to obtain the required information.

Free will can, however, be cast into an alternative formulation in order to render the concept better suited for analysis. Before proceeding to attempt to formulate a definition of this kind, let us temporarily ponder over the characteristics of the problem we want to solve. Imagine a person in a windowless, soundproof room without radio, tv, mobile phone, internet or any other connection to the outside world. We wonder whether the behaviour of this person is in principle predictable for a Laplacian demon that has complete knowledge of all the present physical details of the situation, including the full composition of the person's body and the positions of all its atoms and the forces between them, as well as a full description of the room in which the person is situated. In a physicalist view, what is required is a solution to the physical laws that govern the system at hand. If the demon could succeed with such a task, free will is strongly questioned. The behaviour of the individual would be completely determined by externally identifiable causes, not from any independent first person choices. Clearly, an adequate definition of free will must provide ability to distinguish between the two cases where the demon can predict the individual's behaviour and when it cannot. This is not sufficient, however, for demonstrating free will. Clearly, from the individual's point of view wilfull actions must have been consciously, rather than subconsciously, considered in advance. We will thus discuss whether this requirement can be secured.

Following Carnap (1950), a transformation from the pre-scientific explicandum to a more precise scientific explicatum would have the advantage of rendering free will a concept properly suited for a formal analysis. In this spirit the following definition will be employed in the present work: *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 refer to rational preferences or desires by a cognitive system for future actions. Furthermore, by '*ontologically open system*' is meant a causal high-level system the future of which cannot, even in an *a posteriori* sense, be reduced to the states of its associated low-level-systems, not even if the system is rendered physically closed.

We motivate this definition of free will as follows. Experience tells us that basic *low-level* phenomena, like individual interactions between neurons in the cerebral cortex, are causal and essentially deterministic. Quantum mechanics tells us, however, that certain corrections of a statistical character must be taken into account, as discussed further on. We will assume that account has indeed been taken of the latter effects when we henceforth make use of the term 'deterministic'. If also the *high-level* neuronal functions and processes being associated with consciousness are deterministic in the sense that they are reducible to low-level processes or properties, it may be quite natural to draw the conclusion that expressions of will are governed by processes outside its conscious control. This is a feature of the classical, deterministic argument against free will. On the other hand, behaviour related to ontologically open conscious systems is not directly reducible to earlier physical low-level neural states. As discussed in the next section, this is a consequence of the ontologically emergent properties of consciousness. It should be noted that ontological emergence does not straightforwardly imply ontological openness; even if high-level properties cannot be simply reduced to those of low-level it must be shown how epiphenomenalism is avoided and *how downward causation is possible*.

The concept of 'reduction' is central for the argument. Unfortunately, 'reduction' is a widely debated topic among philosophers and there is limited consensus when it comes to details (van Riel and van Gulick 2018, van Gulick 2001). It is in our view reasonable to assume, as van Riel and van Gulick do, that *ontological reduction* should entail "identification of a specific sort of intrinsic similarity between non-representational objects, such as properties or events". An ontologically irreducible property, if it exists, could not be determined by its low-level-properties or behaviour; it could not be characterised by a statistical or law-like behaviour in relation to its low-level components. It is not implied by nature. It is argued in Scheffel (2019), using arguments from algorithmic information theory and quantum mechanics, that even assuming causality, the extreme complexity of consciousness, in an ontological sense, 'shields' the dynamics of high-level conscious activity from that of its associated low-level components, the neurons. The implication for consciousness is that its high-level properties are not ontologically implied by its low-level neural activity.

Returning to the definition of ontologically open systems, we need to distinguish between open and closed physical systems. 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 states. This becomes clear if we extend the size of the system to also include all of its external influences. Such an extended system may indeed be *physically closed*, causal and deterministic. A physically closed system is not influenced by any processes outside the system itself. We will, in the next section, argue that consciousness ontologically behaves as a physically open system even in situations that could be classified as physically closed. This is indeed what is meant by an ontologically open system.

For the sake of completeness we should, when discussing the dynamics of open and closed systems, account for that quantum mechanics implies that determinism does not fully apply at the very micro-level. The uncertainty principle of quantum mechanics implies that nature is 'blurry' at the sub-atomic and atomic particle levels in the sense that, for example, the simultaneous position and velocity of a particle are quantities that cannot, even ontologically, be assigned exact values. For larger clusters of particles, however, like the molecules that make up the neurons, this effect is of much less importance. The concept of 'adequate', or 'statistical', determinism (Bitsakis 1988, Goldberg 2018) has been coined to emphasize that the statistical determinism of macroscopic processes holds with hich accuracy for systems like neural networks, even if quantum uncertainty may be important on the very micro-scale. Thus, we may say that on the macroscopic level chance is transcended and transformed into necessity (Bitsakis 1988).

Returning to the definition of free will stated above, it is emphasized that the desired actions of a free consciousness must not turn into anything other than intended; behaviour must be consistent with the agent's intentions. By 'intention' we here adhere to the everyday definition 'determination to act in a certain way'. Now, if I wish to consider what to eat for dinner, such a reflection must be possible. My choices and actions must consistently and adequately follow my will. The phrasing '*takes place according to its intentions'* is deliberately somewhat vague in the sense that 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 also that we 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 such a manner that consciousness, in spite of being controlled this way, experiences intentions as its own. So-called 'character decisions', being decisions based on our experiences and consolidated positions that we make without active reflection, we treat in this context as conscious. We will return to these.

There is a subtle, but important, observation to be made. Even if our conscious desires and decisions would be completely ruled by subconsciousness, the latter has, if the combined conscious/subconscious mind constitutes an ontologically open system, capacity for choices that are not predetermined. Thus, even if subconsciousness rules the mind, the individual can be regarded as morally and legally responsible for its activities due to the ontologically open character of its mind. As such it has, over time, had the ability to consciously and subjectively integrate the consequences of its actions into its considerations. Hence the debate concerning to what extent subconsciousness rules our decisions is essentially irrelevant as far as moral and legal matters are concerned if it can be shown that the human mind, or consciousness, behaves as an ontologically open system. The role of subconsciousness for free will is discussed in more detail in section 4.

To sum up, we have cast the characterization of free will as 'the ability to act differently' into an alternative, more scientifically useful formulation in order to improve the methodological conditions to address the free-will problem. The gist of the more traditional definition is retained, but the vague and immeasurable 'act differently' is replaced with the notion of consciousness as ontologically open. If consciousness, even in instances when it may be regarded as a physically closed system, can be shown to be ontologically irreducible, there is room for subjective, willful and unpredictable actions. The task is now to address the, as it seems, inhibiting circumstance that the mind must feature a deterministic character in order to enable coherent low-level thought processes and consistent performance of its intended actions, while simultaneously feature an ontologically open nature in order to permit high-level self-caused actions. If this potential contradiction can be dissolved, there is room for free will. It is here that the ontologically emergent character of consciousness plays an important role. Next we aim to show that the associated ontological irreducibility of consciousness to low-level neural states renders consciousness an ontologically open high-level system.

3 Consciousness, determinism and downward causation

The question thus arises how it would be possible for the brain's essentially deterministic, neural low-level activity to facilitate ontologically open behaviour at the higher inter-neural levels related to consciousness, considering that humans and consciousness are of the physical world.

In the present physicalist approach we will assume that the evolution of a specific isolated system, if repeatedly started from identical initial conditions, will always be identically the same. Surely, this is a statement in support of determinism? The answer is certainly yes if we consider a system governed solely by low-level causal processes. Each link in the chain of cause and effect is in principle identifiable and can be expressed in terms of, or reduced to, laws for physical low-level behaviour. Any completely specified low-level state would uniquely generate subsequent low-level states that are scientifically, or third-person, identifiable. We must of course also assume here that there exist no inherent random processes (like radioactive decay), that could have a causal impact.

If, on the other hand, the system features emergent properties, there will appear links in the causal chain of events that are not reducible to low-level. These are due to more complex high-level properties of the system. Since the low-level chain of cause and effect is broken by causes related to high-level phenomena we may also speak of downward causation. Hence low-level determinism does not characterize the system. We may assume, however, that low- and high-level processes *in combination* preserve our physicalist assumption that the time dynamics of two identical versions of the system, run in parallel, should be the same. But it is a mistake to conclude that the time evolution is *implied* by the initial state since this would assume that low-level cause and effect holds all the way. We will soon discuss this in more detail. List (2019), when discussing conscious volitional processes, speak of indeterminism at the high-level in order to capture the fact that low-level determinism, being the basis for third-person observations and predictions, is put out of play for the system as a whole. As we will now see, the situation for these systems may be compared to that of open physical systems, where external phenomena can have an influence on the dynamics.

Let us return to consider the behaviour of a hypothetical single conscious individual placed in a closed room, without contact with the outside world. We are interested in whether predictions of the individual's behaviour in a certain future time interval are in principle possible. For the sake of argument let us first consider an imagined case that we would deem as fundamentally indeterminable with respect to the individual's choices and actions. If the individual, before making 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 determinable. There is no possibility to predict or explain the actions of this individual; the influence of the genie's advice on the individual's behaviour is comparable to when the dynamics of an open physical system is affected by external influences. Since the genie may affect the individual's choices or decisions, we must infer that the will of this individual is not simply the result of causal and deterministic dependence on its initial low-level set-up and conditions in the physical world. 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 as well as the forces acting upon them, the future of the universe would be in principle determinable. This argument, however, implicitly assumes the continual action of the (low-level) laws of nature. In the thought experiment, the genie has the effect of breaking this chain of events.

Returning to reality, we will now assert that the genie of the thought experiment can, with a similar result, be replaced by the individual's ontologically emergent conscious thought processes, including subjective preferences acquired during the individual's earlier history. Will is about planning; thus experience plays a central role. The individual's experiences are personal and internally rated subjectively, and subsequently stored as memories, constituting a basis for future preferences. These preferences are consciously or unconsciously consulted, similarly as in the case of the genie, when making decisions. Such processes are *ontologically emergent* processes where subjective positive or negative connotations have been related to various events, actions and choices. Thus consciousness acts as an open system in the sense that its current neural activity is ontologically detached from its current physical low-level situation. The fact that one in principle can, atom by atom in a Laplacian sense, construct the individual's entire network of coupled neurons is not relevant here. The system has built in subjective preferences, the character of which are ontologically ’unknown’ (memories have no ontological meaning considered at low-level), featuring an independence comparable to that of taking advice from a genie. Ontological emergence is crucial in that it decouples the physical low-level state of the individual as a system from its subjective properties and behaviour. It grounds freedom rather than lawfulness.

The main point of the genie thought experiment is to introduce *an element which is missing in a third person, or ontological, representation* of the mind. This element is beyond the third person notion of deterministic factors in the dynamics and helps to understand downward causation. Think of it this way. Assume, for the sake of argument, that an emergent property P of a conscious system formally can be found from the time-dependent solution of a set of neurophysiological relations for the system, modelled by the equation $\mathrm{D}\mathbf{f} = \mathbf{0}$, in which $\mathrm{D} = \mathrm{D}(\mathbf{f})$ is a linear or nonlinear time- and space-dependent matrix differential and/or algebraic operator working on the variable vector $\mathbf{f} = \mathbf{f}(t,x,y,z)$ with components f_i ($i = 1\ldots N$) that represent the N functions and properties that formally provide a complete description of the conscious system. Since we assume that P is an emergent property, it is in principle impossible to, in a third-person perspective, specify all the functions f_i in detail. But neurophysiology tells us that reasonably accurate theories (at least in principle) can be constructed for limited subsets of neural interactions related to the realization of the property P, such as firings of clusters of neurons. These theories, associated with a third-person view of cortical neural processes, would necessarily employ a reduced set of variables, say $f_1,f_2,\ldots f_M$, for which $M < N$, since the conscious system features further properties than those associated with low-level. Assuming that the property P is ontologically emergent, the variables $f_{M+1}\ldots f_N$ may have a relation to first-person processes only; P cannot be reduced to a physical, low-level relation to these variables. This means that the variables $f_{M+1}\ldots f_N$ and the subset of system relations $\mathrm{D}\mathbf{f} = \mathbf{0}$, for f_i with $i = M+1\ldots N$, determining their temporal evolution, represent a degree of freedom for consciousness, not deterministically related to low-level, third-person accounts of neural processes. We may view this abstract formalisation as a representation of causal laws for the high-level emergent properties that enable conscious thought. It is the associated degree of freedom that decouples consciousness from low-level determinism and allows for mental processes associated with downward causation. Also MacLaughlin (1992) and Chalmers (2006) discuss the possibility for irreducible high-level phenomena to exert a causal efficacy and open up for the existence of high-level laws.

To elucidate the mechanism of downward causation in this context we may employ the thought experiment introduced in Scheffel (2019). A particular type of human-like robot, equipped with body parts, limbs, joints and muscles, is able to walk and run. It could not, by any means, be *designed* to jump without falling, however, due to its particular construction and its complexity. Furthermore, the robot is designed to store in its memory, and make use of, movements that would be advantageous for the tasks it was programmed to carry out. After having been deployed on an island for a certain time, together with other identical robots, in order to carry out certain duties (all being able to communicate with one another) it was later surprisingly found that the robots had *evolved* the ability to jump without falling. The robots thus carried out new tasks, like reaching new parts of the island that previously were inaccessible due to obstacles like ditches.

In this thought experiment no theory can describe the evolved property to jump. This property is thus *epistemologically* emergent. Had the designers of the robot been asked before returning to the island to theoretically model any specific task to be carried out by the robots, jumping would not be included in their models. We may thus expect that their theories would fail to provide an adequate picture of the robot activities on the island. Any attempt to describe, model, understand, predict or control these robots would be incomplete. Referring to the formal reasoning above, it is clear that the models would employ only a limited number of low-level variables M, being determined from M relations or equations, failing to include the additional degree of freedom available for the jumping robots. The robot's ability to jump is a property, or a variable, that should be included in a complete model of its dynamics. Since we assume causality, we may expect that this additional variable for the dynamics is associated with lawful behaviour (Bunge 2017) that, in principle, can be formalized into at least one additional dynamical equation. We may furthermore assume that this equation should couple to the M low-level equations of motion for the robots, since jumping should be accounted for in order to obtain a complete model of their dynamics. The problem is, of course, that jumping is an emergent property in relation to these particular robots, implying that it is epistemologically impossible to construct the full set of M+1 equations, determining the robot dynamics. The additional degree of freedom for the jumping robots is thus decoupled from, or independent of, these equations and yet real. Its influence is precisely analogous to the mechanism of downward causation. We say ”analogous” here, since downward causation is a phenomenon which belongs to an ontological, rather than an epistemological context. We could, however, perform a similar reasoning as above when discussing the role of downward causation for the dynamics of ontologically emergent phenomena. The difference is that the impossibility to reduce jumping to a theory is substituted with the irreducibility of jumping to low-level properties of the robot.

The emerged property, to be able to jump, was here apparent from inspection, that is from a third-person perspective. Let us now relate this thought experiment to consciousness and free will. Thus we move from epistemological to *ontological* emergence. This implies, as we have discussed, a higher degree of complexity; a sufficiently complex system could develop ontologically emergent properties. In the example of the Jumping robot, this would mean that its evolved ability to jump would be irreducible to its low-level properties, even if the computational capacity of the universe was available. This would be the case when, for example, the positions and motions of all its limbs must be tailored with a very high degree of precision. Hence we could carry out a similar discussion as above for the case that jumping, as a property of the robots, evolves as an ontologically emergent property that can not be deterministically accounted for, not even in principle. It is, of course, not likely that the robots will develop such behaviour but we are now able to see how a similar case can be argued for consciousness and will. The brain, with its extremely complex cortical neural network, may in a similar manner feature properties that cannot, neither epistemically nor ontologically, be deterministically reduced to low-level neuronal properties and processes. Consciousness would, in analogy with the Jumping robot, feature a degree of freedom that is beyond deterministic processes at the physical low-level, allowing for downward causation. Whereas the robot's ability to jump was distinguishable in a third-person perspective, the activity of consciousness and will is, however, distinguishable from a first-person perspective only. The standard, third person, scientific and low-level deterministic relation to consciousness halts as emergent behaviour takes over. It cannot reach over this barrier to represent and contribute to understanding of subjective first person experience.

As discussed at the beginning of this section, there is indeed reason to assert that determinism, in a standard interpretation, is an *inadequate* concept for fully characterizing the causal situation for mental processes. Speaking of determinism in relation to neurophysiological processes, we ususally refer to physical, *low-level* determinism, at the atomic and molecular levels. As we have argued, high-level mental processes are also dependent on emergent properties, associated with complex large scale phenomena. Again considering the Jumping robot we may assume, for the sake of argument, that its complexity renders jumping an ontologically emergent property. The robot may consider jumping instead of walking or running for achieving certain tasks. Jumping is a high-level process and as such autonomous with regards to low-level determinism since its behaviour is irreducible to low-level. This is analogous to situations when we make conscious decisions based on our preferences and previous experiences. Hence standard, low-level determinism cannot represent mental will processes. This completes our argument for that consciousness is an *ontologically open* system. Conscious activity cannot, even in an *a posteriori* sense, be reduced to the states of its associated low-level-systems, not even if the conscious system is rendered physically closed.

List (2019) argues that, instead of low-level physical determinism, mental processes are governed by ”agential indeterminism”. Semantically, this is an unfortunate label, since it leads the thought to probabilistic, or random, processes. Furthermore the argument for agential indeterminism fails to consider that, although what he terms ”psychological-level states” may have been caused by non-identical low-level neuronal states and thus on a psychological-level may feature bifurcating futures, it must be explained why and how psychological-level states become, at least partly, independent of physical (low-level) states. Low-level determinism does not vanish by mere focusing on high-level processes. In the present work we have introduced the concept of ontological openness in order to show how ontological emergence decouples standard, low-level determinism from emergent high-level conscious activity, thus enabling downward causation.

We started this section by asserting that causality implies that the evolution of a specific system, if repeatedly started from identical initial conditions, will always be identically the same. Have we now resolved the resulting difficulty for free will? Central for our argument is ontological openness, exemplified by the genie. Clearly, if the relation to the genie is external to the physical low-level system on which the specific consciousness supervenes, expressions of will are autonomous in relation to the physical low-level system. If the genie, on the other hand, is internal to the system, it would appear that the physical low-level situation determines its actions and we are back to the original problem. As we have seen, this is however not the case, due to downward causation. This high-level process defeats low-level determinism and may be associated with causal laws at high-level. It is the genie, or rather the equivalent ontologically open processes related to consciousness, that enable high-level downward causation. By taking into account subjective, high-level first-person experiences and memories, autonomous will is formed.

In short, the problem with claiming that determinism implies that the evolution of a specific system, if repeatedly started from identical initial conditions, will always be identically the same in relation to consciousness is the fact that although the outcome of a will process can only be a specific one, this does not mean that the outcome is settled beforehand. It is only when indeterminable high-level deterministic processes including accounts of subjective first-person preferences and experiences, associated with downward causation, are included that the evolution of the system becomes fully specified. Low-level determinism does not alone determine the evolution of the system.

To sum up, we have argued that consciousness is an ontologically open high-level system and thus third-person, or ontologically, indeterminable and uncontrollable in principle. Conscious will is, rather than being determined by low-level neural properties, the result of ontologically emergent high-level processes including accumulated subjective experiences, stored as memories. Having eliminated straightforward dependence on low-level neural properties, we have thus also eliminated *epiphenomenalism*.

In the process, we have also discussed how downward causation (Campbell, 1974 and Kim, 2006) enters. We have presented a solution to the question of *overdetermination* with regards to the causal situation for consciousness, termed the causal exclusion principle by Kim (2006). Kim argues that if the dynamics of consciousness is determined by its current state and the laws of nature, then emergent phenomena cannot exist independently; they must be a result of the complete set of conditions already provided. Otherwise we seem to be facing an overdetermined problem. But we can now see that the solution to this dilemma is that emergent properties are of the same nature as the new conditions that may present themselves when a closed system is transformed into an open system. Hence they are *additional* conditions, being governed by associated additional relations. Mathematically speaking, just as many new equations are added as new variables. Thereby overdetermination is avoided. In our example of the person being placed 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 explains how downward causation can take place. Interacting emergent phenomena can determine the development of the system (in this case, the mind) to a large extent independently of the causal situation at lower levels. The nature of consciousness as an ontologically open system removes supervenient bottom-up determinism.

Ontological emergence of consciousness is essential for free will. If consciousness were merely *epistemologically* emergent, an imagined powerful Laplacian demon, with access to all physical information in the universe including all details of the individual's consciousness, could in principle manipulate the individual to act in any specific way by engineering its low-level neurons. An ontologically emergent consciousness is, however, without reach for the demon; it is free in the sense that its action cannot be determined, understood or controlled, not even in principle.

4 Willed intentions and the role of subconsciousness

Free will requires, in line with the definition employed here, 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 decide to return to the pavement in order to avoid an approaching car, she returns. Does she want to make herself a cup of coffee, she makes it. Exceptions that can be identified, such as in the latter case a shortage of coffee or an interruption due to a ringing phone, are not about principal mental limitations but of properties of the outside world.

So far, we have presented arguments for that consciousness/subconsciousness as a combined system meets the requirements for free will. But few would regard this as sufficient; if our volitional decisions, in spite of their ontologically open origin, are unconsciously dictated to us it would be difficult to speak of free will. There is evidence that consciousness in a vast number of situations exerts its will without significant influence from mind processes that we would refer to as subconscious. It should be noted, however, 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 symbiosis between consciousness and subconsciousness; we experience ourselves alternating between actively reacting to the current traffic situation as well as being deeply immersed in our own thoughts. Participation in an intense discussion, where rapid reponse is required, 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 has been 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), who used an electroencephalogram (EEG) and placed ~~small~~ electrodes at various points on the scalp of subjects to measure neuronal activity in the cortex. He found that EEG signals, related to certain wilfull actions, could be recorded as long as half a second before the subjects admitted to having made a decision. Experiments in this field has, however, many possible sources of error, thus criticism comes from several places (Klemm, 2010 and 2016, Baumeister et al, 2011). We now briefly consider some of these arguments.

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 a car 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, find the time required 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 likely have provided us with a mechanism that automatically disconnects consciousness in favour of subconsciousness, like 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. In many similar situations, however, subconsciousness cannot reasonably play a significant role.

The continuous cooperation between consciousness and the unconscious points to a second argument why consciousness is not 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 supports the idea that also subconscious neural processes are ontologically emergent. Thus, whereas determinable low-level processes are associated with communication between consciousness and the unconscious, these systems can both, on high-level, be assumed to behave as ontologically open systems that to a large extent act independently. As pointed out, experience shows that we can consciously cancel impulsive intentions, using ”free won't” (Libet, 1985).

From one perspective, we do not necessarily need to distinguish between consciousness and subconsciousness as separated global systems. Already individual neurological *subsystems* associated with the mind appear to be sufficiently complex to render their interaction ontologically emergent and thus ontologically open. 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 definitions of subconscious and conscious choices is what might be called 'character decisions' (Danto and Morgenbesser, 1957). Based on previous experience and reflections, people accumulate different, often conscious, 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. Clearly, most of us would admit to struggling with some undesirable traits of character, but this fact is not central for the question of free will. 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. The topic is interesting and should preferably be treated in more detail.

Our feelings, thoughts and choices do not simply happen to us. They develop emergently in a cooperation between high-level consciousness and the unconscious. But how, then, can our thoughts and subjective feelings take form in a structured and coherent way? What is the detailed interplay between consciousness and subconsciousness? These important questions are not analyzed here. Of prime interest for free will is that high-level thoughts, subjective feelings and conscious choices arise in a manner which is irreducible and indeterminable in principle.

5 Discussion

The present analysis of free will is consistent with non-reductive physicalism, where mental states supervene on physical states but cannot be reduced to them. Thus there are similarities with Davidson’s theory of anomalous monism (Davidson, 1970) in which the Anomalism Principle implies that there are no strict laws on the basis of which mental events can be predicted or explained by other events. The present work provides an explanation for the non-existence of such laws.

It is of interest to discuss the relation to *naturalistic dualism* (Chalmers, 2007). In this nonreductive theory, with some characteristics common to property dualism, it is argued that there is an unbridgeable explanatory gap between objective and subjective experience. Consciousness is here a fundamental property, ontologically autonomous of the physical properties upon which it supervenes (see also Chalmers 1995). A theory for consciousness would thus call for a set of high-level ”psychophysical laws”, much like electromagnetism requires Maxwell’s equations for a description rather than merely basic Newtonian laws. Although similarities exist with the present theory, it should be noted that the assumed supervenience on a low-level, neurophysiological basis of the present theory leads to a monistic view on consciousness. As ontologically open, the mind features a freedom much like Gödel-unprovable statements do in mathematics. Gödel-unprovable ’high-level’ statements ’supervene’ on provable theorems of standard, ’low-level’ mathematics. Additional high-level Gödel-unprovable statements can be generated by combining Gödel-unprovable statements with themselves or standard mathematics. Complexity at the low-level is the root of all this; it provides independent and unprovable statements in mathematics as well as independence and freedom for the mind in the physical world. But complexity also works at low-level, hence in the theory outlined here both physical and mental properties, supervening on physical substance, interact simultaneously. It would thus seem more natural to associate consciousness with a monistic rather than dualistic view.

We may ask: is the degree of freedom for the will of the present theory consistent with the common characterization of free will as ”able to act differently”? The answer depends on the interpretation of the vague formulation ”differently”. If ”differently” refers to the low-level neurological states on which a mind supervenes, the answer is positive. The details of conscious activity are not implied by low-level neurological states. If ”differently”, on the other hand, refers to high-level conscious considerations the answer is negative. Even a freely (with respect to low-level) chosing consciousness should, by causality, make the same choices over and over again if the neurological basis on which it supervenes is identical for all cases. If it did not its action would be random, being inconsistent with experience.

But isn't this the same as to say that the system is deterministic? Causal closure tacitly assumes that, when consciousness is regarded as a physically isolated system, all potential causes for its future dynamics are contained within the system. As we discussed in the introductory section, the implicit view of the natural sciences is usually that such a system is deterministic in some sense. Given our vast experience of low-level determinism, it might be expected that this sence is low- to high- level determinism. This was, however, found to be erroneous in light of the low- to high-level ontological indeterminacy that follows from ontological emergence. Instead, we note that there is also a high-level base of potential causes for the dynamics of the system. These are indeterminable from, or irreducible to, low-level. Even if complete access to a brain was given for scientific analysis, there would be no chances of, for example, identifying what would be this individual's precise choices of food and wine when selecting from a menu at a restaurant. The situation can be further elucidated by again considering the imagined genie of section 3. For any low-level-specified state of the conscious system, we may ask what will be the genie's advice. We will find that the response will differ for different low-level scenarios, but it is in principle impossible to control the response by manipulating the low-level states. The genie's different advices will always appear as surprises. Causal closure, however, requires the genie's advice to be the same for identical scenarios. But it would not be meaningful to say that it is *implied* by each low-level state; the response is, as we have seen, both epistemologically and ontologically irreducible to low-level.

As argued by List (2019), whether determinism is at hand or not is level-dependent. In the perspective of consciousness as a closed physical system, we have found that low- to high-level indeterminacy renders consciousness ontologically open. This is why it would be misleading to lable a physically closed conscious system (brain in a vat) deterministic. The assumption of causal closure indeed guarantees that all possible causes for its future dynamics are contained in the system, but low-to high-level indeterminism renders the action of consciousness ontologically, or third-person, irreducible to previous physical low-level states of the system. Determinism, in the Laplacian sense, is not satisfied.

Given the indeterminacy of consciousness from low-level, the long standing problem of compatibilism versus noncompatibilism becomes of less interest. Furthermore, the issue of compatibilism loses interest at the conscious high-level also, because activity at this level depends, as we have seen, partly on indeterministic interaction with low-level processes. Thus the freedom granted by low- to high-level indeterminism renders consciousness associated with subjective high-level properties and phenomena such as thoughts, ideas, feelings and remembrances, all contributing to downward causation, the main characteristic of free will.

Then, what is there more to say about *epiphenomenalism*, the notion that mental states are only by-products of the physical states and unable to causally influence these? As discussed above, the form of non-reductive physicalism assumed here is *not* a form of property dualism, or dualism of any sort (see Robinson 2017 for an excellent review). Although mental states are not deducible from basic neurological states, they correspond to physical states; they supervene on these. Non-reductionism follows because of the emergent character of mental states, not because of lack of two-way interaction between physical and mental states. Clearly, epiphenomenalism is inconsistent with the present theory.

We have in this study provided an outline for how the volitional processes of a conscious agent, interpreted as an ontologically open system, can be associated with a large degree of volitional freedom. Since arguments for the main conclusion has been in focus, some of the topics touched upon require further analysis in order to establish more conclusive arguments.

6 Conclusion

It is found that high-level cognitive processes are ontologically open, even though underlying physical laws and low-level neural processes may be assumed essentially deterministic in a standard sense. By an 'ontologically open' system we mean a causal high-level system, the future of which cannot be reduced to the states of its associated low-level-systems, not even in situations where the system is physically closed. The analysis builds on consciousness as an ontologically emergent property of the brain. Thus the activity of consciousness is not determinable, not even in principle. To consider the impact on volitional processes, a methodologically more applicable definition of free will than the widely assumed 'ability to act differently' is suggested. The three associated requirements for free will are all argued to be 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's mind constitutes an ontologically open system. Thus the will, as defined here, is free.

References

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

Bitsakis, E. (1988). Quantum Statistical Determinism. *Foundations of Physics*, 18, No. 3, 331-355.

Bonilla, J. Z. (2019). Why emergent levels will not save free will. *MappingIgnorance*, https://mappingignorance.org/2019/09/30/why-emergent-levels-will-not-save-free-will-1/ and https://mappingignorance.org/2019/10/21/why-emergent-levels-will-not-save-free-will-2/

Bunge, M. (2017). *Causality and Modern Science*. Routledge.

Campbell, D. T. (1974) Downward causation in hierarchically organised biological systems. In Francisco Jose Ayala and Theodosius Dobzhansky (Eds.), *Studies in the philosophy of biology: Reduction and related problems*, pp. 179–186. London/Basingstoke: Macmillan.

Campbell, R. J. & Bickhard, M. H. (2011). Axiomathes 21, 33–56.

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

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

Chalmers DJ (1995) Facing up to the problem of consciousness. Journal of Consciousness Studies 2:200–219.

Chalmers D. J. (2006) Strong and weak emergence. In: P. Davies & P. Clayton (eds.), The Re-Emergence of Emergence: The Emergentist Hypothesis From Science to Religion. Oxford University Press

Chalmers, D. (2007). Naturalistic Dualism. *The Blackwell Companion to Consciousness*, Schneider, S. and Velmans, M. (editors), 359-368. Wiley Online Library.

Danto, A. C. & Morgenbesser, S. (1957). Character and Free Will. *Journal of Philosophy*, 54, No 16, 493-505.

Davidson, D., (1970). Mental Events, reprinted in *Essays on Actions and Events*. Oxford: Clarendon Press, 1980.

Davies, P. C. W. (2004). Emergent biological principles and the computational properties of the universe. *Complexity* 10,11-15.

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

Dennett, D. C. (1997). *Elbow Room*. Bradford Books, The MIT Press.

Goldberg, G. (2018). Temporal Naturalism, Free Will, and the Cartesian Myth: Time Is NOT Illusory and We Are NOT 'Talking Heads', *AJOB Neuroscience*, 9(1): 1–4, 2018.

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

Kim, J. (2006). Emergence: Core ideas and issues. *Synthese*, 151, 547-559.

Kim, J. (2011). *Philosophy of Mind.* Westview Press.

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

Klemm, W. R. (2016). *Making a Scientific Case for Conscious Agency and Free Will*. Elsevier Inc.

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

List, C. (2014). Free will, determinism, and the possibility of doing otherwise. *Noûs*, 48(1), 156-178.
List, C. (2019). *Why Free Will is Real*. Mass.: Harvard University Press.
Lloyd, S. (2002). Computational Capacity of the Universe. *Physical Review Letters*, 88, 237901-1-4.
Macdonald, G. (2007). Emergence and Causal Powers. *Erkenntnis*, 67, 239-253.
MacLaughlin, B. (1992). 'The Rise and Fall of British Emergentism'. In A. Beckermann, H. Flohr, and J. Kim (eds), *Emergence or Reduction: Essays on the Prospects of Nonreductive Physicalism*. New York and Berlin: W. de Gruyter, 49–93.
O'Connor, T. & Franklin, C., "Free Will", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/freewill/>.
Popper, K. and Eccles, J. (1977). *The Self and its Brain*. New York: Springer.
Robinson, Howard, "Dualism", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/dualism/>.
Scheffel, J. (2019). On the Unsolvability of the Mind-Body Problem, *Axiomathes*, 1-22 *https://doi.org/10.1007/s10516-019-09454-x*.
Stephan, A. (2010). An Emergentist's Perspective on the Problem of Free Will. In G. Macdonald and C. Macdonald (eds), *Emergence in Mind*. Oxford University Press, 222-239.
van Gulick, R., Reduction, Emergence and Other Recent Options on the Mind/Body Problem: A Philosophic Overview, *Journal of Consciousness Studies* 8: 1–34, 2001.
van Riel, R. & Van Gulick, R., Scientific Reduction, *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/scientific-reduction/.
Wegner, D. M. (2002). *The Illusion of Conscious Will*. Bradford Books, The MIT Press.
Weissman, D. (2019). *Metaphilosophy*, 50(5), 743-747.
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.