Free Will of an Ontologically Open Mind

Abstract

The problem of free will has persistently resisted a solution throughout centuries. There is reason to believe that new elements need to be included in 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 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 dynamics of which cannot be reduced to the states of its associated low-level neural systems, not even if it is rendered physically closed. A positive answer to the question of free will is subsequently outlined.

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. We will argue that one reason for the problematic situation can be traced to the common notion of free will as ’the ability to act differently’ or that we ’could have done otherwise’. For example, why should a freely acting agent have reason to behave differently in two identical situations? In the next section, we will suggest an alternative definition of free will.

It has been reasoned that consciousness cannot be represented by a reductionistic theory and, as a consequence, that the mind-body problem is unsolvable (Scheffel 2020). The associated *epistemological* emergence of consciousness is of interest for the problem of free will since if, on the other hand, a detailed theory for consciousness could be designed, then its behaviour would in principle be computable or could be simulated. Thus, if we could understand consciousness reductionistically there would be little room for free will, a consequence that has not received much attention in the literature. It was furthermore found that consciousness, as a high-level property of the mind, is *ontologically* emergent with respect to its low-level neural states. A high-level property was defined as 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.

The assertion that extremely complex systems may feature ontologically emergent properties 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). 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 the available

physical states of nature are insufficient to accommodate all required information associated with a formal *representation* of the system's properties - then the system features ontologically emergent properties. Thus, although consciousness supervenes on low-level neurobiological states, it was found that consciousness is not ontologically reducible to the properties of these because of the extreme complexity of the cortical neural network.

In the present work, we will contend 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.

A physicalist 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 and the forces that will act upon it. We will touch upon microscopical uncertainties caused by quantum mechanical effects later on.

Causality and determinism are essential elements in relation to free will. These concepts can be interpreted by considering the order of related events in time. Causality is *a posteriori* in the sense that it, by definition, entails 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 we traditionally term it deterministic. By this we mean that any transition from one state to the next is fully determined by the initial state. As we will see in the following, however, the evolution of physically closed systems featuring ontologically emergent properties are usually indeterministic in this classical sense, on the grounds that there will exist causal transitions between high-level states that are principally irreducible to low-level states due to downward causation. This is a central distinction used in the present work.

Determinism, corrected for quantum mechanical uncertainty, is usually implicitly assumed in the physical sciences; in principle it enables prediction through the use of theories, like natural laws or simulations. An *open* physical system may however, by definition, interact with the world external to the system (Ismael, 2016). While preserving causality, it cannot be assumed deterministic. Causality does not imply determinism since it does not require particular, individual causes to uniquely specify the future of a system. In the present physicalist approach it is assumed that causal closure holds; no physical event has a cause outside the physical world.

Are similar positions on free will found in the previous, vast literature? Since we argue that emergence is a required element of a solution to the problem, the number of related publications 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 explicit reference to emergentism is avoided, 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 "Why free will is real" (2019), an extensive literature study has been carried out. 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 (Murphy 2009, Campbell and Bickhard 2011) and arguments for causal efficacy without downward causation (Macdonald 2007).

Also 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 weak (Weissman 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) requires further support. 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. The main line of thought and outline of the paper is as follows. Free will requires a definition in the spirit of Carnap, that is its characterization should be similar to our everyday notion of the concept as well as exact, fruitful and simple. This is the topic for section 2. In section 3 we address whether consciousness enables sufficient conditions for free will. Of particular importance is downward causation, which in turn assumes ontological emergence of high-level cognitive processes. It must now be clarified how emergence facilitates deterministic independence of low-level neural processes. Thus the concept of ’ontologically open’ systems is introduced. These 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 they are physically closed. 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. It is then asserted that consciousness satisfies all three conditions for free will. In section 4 the role of subconsciousness is considered. Discussion and conclusions, finally, can be found in sections 5 and 6.

2 Definition of free will

Common characterizations of free will like, for example, ’ability to act differently’ or ’could have done otherwise’ pose problems. How would we resolve the question whether consciousness has an ’ability to act differently’? What information is to be found? In this characterization ’differently’ is about outcomes, which in principle can be identified experimentally. But ’differently’ also refers to the neural processes that are involved in the agent’s deliberation. These could be of strictly deterministic, low-level origin or be associated with emergent high-level, conscious considerations that facilitate downward causation. As will be discussed, the degree of freedom for the will is quite different for these two cases. In both, however, it may well be argued that there is no reason for a conscious agent to behave differently in identical situations. Let us, for the sake of argument, suppose that two identical isolated conscious systems have been designed. Monitoring their time evolution, we would certainly not expect these to show divergent behaviour. Moreover, ’ability to act’ concerns a cognitive and subjective first person process to which we have no third person access, neither theoretically nor experimentally.

Free will can, however, be cast into an alternative formulation in order to render the concept better suited for analysis. Before formulating 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; wilfull actions should not have been subconsciously generated.

Following Carnap, a transformation from the pre-scientific notion of free will to a more precise scientific explicatum can be made with the method of explication (Carnap, 1950). 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 as follows. Experience has shown 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 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 widely debated 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 employ the definition mentioned in the Introduction, by van Riel and van Gulick. An ontologically irreducible property, cannot be determined by its low-level-properties or behaviour; (it cannot be characterised by a statistical or law-like behaviour in relation to its low-level components.) It is not implied by nature. In Scheffel (2020) algorithmic information theory and quantum mechanics are used for arguing 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.

In order to specify ontologically open systems, we need to distinguish between open and closed physical systems. Phenomena relating to classical *open physical systems* are generally causal, but indeterministic. 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 now extend the size of the system to also include all of its external influences. Such an extended, classical system contains all of its causes and thus constitutes a *physically closed*, causal and deterministic system. No processes outside the system itself can have any influence. We will, in the next section, however argue that consciousness has features of an open system even though the system's low-level basis is 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 shows 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 basic 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 adhere to the everyday definition 'determination to act in a certain way', enabling the agent to have control. 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, point to be made. Even if our conscious thoughts, 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. In consequence, the individual can be regarded as morally and legally responsible for any associated activity. It has, over time, had the ability to consciously and subjectively integrate the consequences of its actions into its considerations. The debate concerning to what extent subconsciousness influences our

decisions is thus less relevant in relation to moral and legal issues if it can be shown that the human mind, or consciousness, features ontologically open properties. 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, scientifically more useful formulation in order to improve the methodological conditions to address the free-will problem. The gist of traditional definitions is retained, but the vague and undecidable phrase '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 independent 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. It is indeed here that the ontologically emergent character of consciousness plays an important role.

3 Consciousness, determinism and downward causation

When discussing conscious volitional processes, indeterminism in relation to high-level can be argued for, since low-level determinism, being the basis for third-person observations and predictions, may be put out of play for the system as a whole (List, 2019). As we will now see, the situation for these systems may be compared to that of open physical systems, where external phenomena can influence the dynamics.

Let us again 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 indeterministic at the neural level 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 deterministic at low-level. 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 external effects on an open physical system. 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 deterministically given. 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. In these ontologically emergent processes 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', or unrepresentable (memories have no ontological meaning considered at low-level), featuring an independence comparable to that of taking advice from a genie. Ontological emergence here plays a crucial role 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. We can now see that what is essential for the argument is not that the genie is external in any sense, but rather that it features an independency in relation to the conscious agent.

The main point of the genie thought experiment is thus 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. We may think of it this way. Assume, for the sake of reasoning, that an emergent property P of a conscious system formally can be found from the time-dependent solution of a set of neurophysiological relations, 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 $\mathbf{f} = \mathbf{f}(t,\mathbf{x},\mathbf{v})$, with space and velocity vectors $\mathbf{x} = (x,y,z)$ and $\mathbf{v} = (v_x,v_y,v_z)$, having 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 directly associated with low-level. Assuming that the property P is ontologically emergent, the variables $f_{M+1}\ldots f_N$ relate to processes on first person level 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 $i = M+1\ldots N$, that imply their temporal evolution, represent a degree of freedom for consciousness, not deterministically related to low-level, third-person accounts of neural processes. This abstract formalisation can be seen as a representation of causal laws for the high-level emergent properties that enable conscious processes. The associated degree of freedom 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 above, and the mechanism of downward causation, the thought experiment introduced in (Scheffel 2020) is instructive. A particular type of human-like robots, equipped with body parts, limbs, joints and muscles, are able to walk and run. They could not, by any means, be *designed* to jump without falling, however, due to their particular construction and its complexity. Furthermore, the robots are designed to store in their memory, and make use of, movements that would be advantageous for the tasks they were programmed to carry out. After that a number of identical robots (all being able to communicate with one another) were deployed on an island for a time in order to carry out certain duties, it was later surprisingly found that the robots had *evolved* the ability to jump without falling. The robots thus could carry out new tasks, like reaching 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. Hence 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, found 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 mechanistic laws (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. 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, describing 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, carry out 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 entire computational capacity of the universe were 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. We could thus argue similarly as above for the case that jumping, being a property of the robots, evolves as an ontologically emergent property that cannot 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, in a similar manner features properties that cannot, neither epistemically nor ontologically, be deterministically reduced to low-level neuronal properties and processes. Consciousness, in analogy with the Jumping Robot, features degrees of freedom that are 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 usually refer to physical, *low-level* determinism, at the atomic, molecular and cellular levels. As we have found, high-level mental processes are

also dependent on emergent properties, associated with complex large scale phenomena. Whereas the associated processes at the emergent high-level feature law-like behaviour (like for the Jumping Robot's high-level ability to jump), that in a sense may be termed deterministic, it is the partial independence of low-level processes, granted by ontological emergence, that removes classical determinism for the system as a whole.

In summary, we have reasoned above that consciousness is ontologically open, primarily as a result of its ontologically emergent character. 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.

Having argued against that the activity of a conscious mind is reducable to its low-level basis, there still remains the possibility of *determinism at high-level*, seemingly conflicting with free will. For example, identical brains-in-vats could then be initiated from a large number of initial conditions whereafter they could be observed in order to establish an empirical theory for their dynamics. Hence predictability and no room for free will. This reasoning is, however, erroneous since the variables that are ontologically accessible at low-level, to be included in any attempt for an empirical theory, are insufficient to describe the high-level dynamics, even if it were deterministic on this level. The conscious agent (brain-in-a-vat) acts, as we have described above, also in relation to emergent and inaccessible high-level properties or variables like, for example, subjective memories. Incidentally it may be remarked that this poses constraints on the accessible levels of understanding in research fields like psychology, sociology and economy, where human activity plays a central role.

Still, the laws of physics and causal closure imply that two identical, closed physical systems must feature identical time dynamics, with possible deviations only related to quantum mechanical uncertainty. This is also what is required from supervenience; any difference at high-level must have its root in a difference at low-level. Would not this circumstance, even in the present context, contradict the possibility of free will? We may recall that the criticism put forth in section 2 against the characterization of free will as 'ability to act differently' mainly focuses on that there is no reason for two identical, isolated conscious systems, or agents, to behave differently, independently of whether they can make free choices or not. Identical time dynamics for the systems does not exclude free will.

This can be understood by again employing the thought experiment of the Jumping Robot. Let us assume that the robots have been programmed to carry out all assigned duties as efficient as possible. Now one of the robots finds itself close to a stream, with the nearest bridge crossing it at an appreciable distance. The task at hand requires the robot to cross the stream. In a low-level perspective, in which the emergent process of jumping cannot be integrated, the only option for the robot is to head for the bridge. In a high-level perspective, the robot is most likely to gain time by jumping over the stream, either directly or by taking a few steps back in order to attempt the jump running. If jumping as a property were ontologically emergent, as for processes involving conscious will, the robot's decision to jump is independent of processes at low-level by virtue of ontological openness. It involves the action of laws and variables at high-level, as discussed earlier. It is a case of downward causation, consistent with its supervenience on low-level states. At the same time, full low-level deterministic control is put out of play. Thus, if a number of scenarios, identical to this, could be facilitated the robot would solve the situation exactly the same every time, say by first taking a few steps back. This is a choice, irreducible to the robot's physical low-level basis. We can here see why it is not fruitful to relate free will to an 'ability to act differently'.

A comparison can be made with Gödel non-decidable propositions in mathematics. Formal systems in which a reasonable amount of elementary arithmetic operations can be carried out can express propositions, the truth values of which are independent of the axioms of the formal system - they are "emergent" with respect to what can be expressed within the system. The phenomenon of independence, or unrepresentability, is not novel; it occurs also in physics. Richardson (1968) has proven that the theory of elementary functions in classical analysis is undecidable. Thus the answers to a host of problems in classical mechanics are independent of the axioms of mechanics, just like the parallel postulate of Euclid is independent of the remaining axioms of plane geometry, and cannot be deduced from them. Particular examples of physical systems that are unpredictable at high-level, with properties that are undecidable, have also been put forth (Moore 1990, da Costa och Doria 1991, Cubitt et al 2015, Ippolito and Caprara 2021). Pitowsky (1996) points to relatively simple physical systems with properties being impossible to compute in any representational system. Wolfram (1985) finds that the most efficient procedure for determining the future of many physical systems is not by computation, but by their own evolution (computational irreducibility).

To sum up, we have argued that consciousness is an ontologically open high-level system and thus third-person, or ontologically, irreducible. Conscious will is, rather than being determined by low-level neural properties, the result of ontologically emergent high-level processes including accumulated subjective experiences in the form of 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 may now address 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.

Kim's argument is however flawed on the grounds that he assumes that a complete set of laws and rules are provided for handling all possible properties and processes of the system. This may not be the case. Euclidean geometry, for example, defined without the fifth postulate (the parallel axiom) is sufficient for deriving a number of axioms and results in geometry. But without a fifth axiom, calculations of angles in a triangle becomes impossible since their sum depends on the precise formulation of this axiom. Similarly, emergent new properties require additional laws, or axioms, in order to make the system complete with respect to the dynamics of these properties. 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. This is also found by considering the analogy of the Jumping Robot, for which there cannot exist low-level theories for, say, the length L of its jump as function of basic parameters; jumping could not be predicted at low-level. At high-level, jumping is an accessible property and the existence of a relationship for L in terms of parameters such as the robot's configuration, its speed and the character of the ground can be assumed in principle. This corresponds to one more 'equation' at high-level for the 'variable' L. If jumping were an ontologically emergent property, it would be an instance of downward causation. We emphasize again that since high-level determinism, seen from a low-level perspective, is not a meaningful concept the Jumping Robot merely serves as an analogy. Emergent properties have, as far as deterministic control is concerned, the same impact on the evolution of the system as external influences have on an open

system. We have thus removed the problem of overdetermination and shown how downward causation can take place. Interacting emergent phenomena can specify 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 specific ways by engineering its low-level neurons. An ontologically emergent consciousness is, however, without reach for the Laplacian 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 causal 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 good 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 partly been 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 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 deterministic 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 predominantly to be conscious rather than unconscious.

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 indeterministic as seen from low-level.

5 Discussion

The theory of free will, being outlined here, 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. We have found that, as ontologically open, the mind features a freedom much like Gödel-unprovable statements do in mathematics. Gödel-unprovable 'high-level' statements 'supervene' on (are formulated from) 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 high-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 present theory both physical and mental properties, supervening on physical substance, interact simultaneously. It is thus, in this sence, more natural to associate consciousness with a monistic rather than a dualistic view.

We may ask: to what extent is the degree of freedom for the will as outlined here consistent with the characterization of free will as the 'ability 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 again positive. As we have shown above, ontologically emergent high-level activity is beyond low-level causal laws. We have, however, also found that a characterization of free will as 'ability to act differently' is unfruitful.

The following question naturally comes to mind: constructing the Jumping Robot atom by atom, at what point do the laws of physics for its low-level basis cease to apply also for its high-level dynamics? If it is the case that downward causation may also cause surprises in the robot's dynamics in relation to what we expect from measurements and calculation, this needs to be resolved.

The answer is that if matter is arranged in certain ways, ontologically emergent systems can arise. These have properties, the dynamics of which can neither be calculated or be expected from, nor reduced to, the properties of the associated low-level systems. Similarly as when we, with the 'matter' of mathematics, construct statements with truth values that standard mathematics cannot decide, certain material systems have properties that are independent of their low-level constituents. This is shown by the Jumping Robot example. The properties are not deterministic in the ordinary sense (in relation to the properties of the low-level states) thus they are also unrepresentable. This does not apply to all complex material systems; the requirement is that ontological emergence comes into play. As a comparison,

only a limited subset of all mathematical propositions feature a truth content being independent of standard mathematics. Hence traditional physical laws cease to apply in full as soon as the associated matter has been arranged in such a special way that ontologically emergent properties arise. Consequently the conditions and dynamics at ontologically emergent high-level is beyond low-level causal laws.

We have found that low- to high-level indeterminism renders consciousness ontologically open. This is why it would be misleading to lable even a physically closed conscious system (brain in a vat) deterministic even though, as discussed previously, high-level determinism is expected. 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 makes 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. This does not imply that physicalist freedom of the will is equivalent to that of a dualistic world in which the soul is, per definition, to a large extent independent of the physical. Rather, we have seen that causal closure in a monist world is fully compatible with the basic characteristics of our notion of free will.

Given the irreducibility of consciousness to low-level, the problem of compatibilism versus incompatibilism is of less interest in this context, and is consequently not discussed. Our argument that the mind cannot be deterministically reducible to low-level is strictly not libertarianism (Ginet 1989, McCann 1998), since we do not claim that standard low-level determinism is false. Neither is it meaningful to characterize our theory as compatibilistic because it is not sufficient that low-level determinism is compatible with free will. The freedom granted by low- to high-level indeterminism renders consciousness associated with subjective high-level properties and activities such as thoughts, ideas, feelings and remembrances, all contributing to downward causation, a main characteristic of free will.

The presence of downward causation in neural activity implies that the theory outlined here is scientifically falsifiable. In the event that subjective conscious experiences would be fully reducible to, or explainable by low-level neural activity, downward causation is ruled out, contradicting the theory.

Thus 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 freedom. Due to the limited space of this article, arguments for the main conclusion has been in focus; some of the topics touched upon in this section will be discussed in detail in later work.

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. Due to downward causation, the activity of consciousness is not low-level deterministic. 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.

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*. Chicago University 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.

da Costa, N. C. A. and Doria, F. A. (1991). Undecidability and Incompleteness in Classical Mechanics, *International Journal of Theoretical Physics*, 30(8), 1041-1073.

Cubitt, T. S., Perez-Garcia, D. and Wolf, M. M. (2015). Undecidability of the Spectral Gap. *Nature* 528, 207-211.

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.

Ginet, C., (1989). Reasons Explanations of Action: An Incompatibilist Account. *Philosophical Perspectives*, 3: 17–46.

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.

Ippolito, E. and Caprara, S. (2021). Undecidability of the Spectral Gap: An Epistemological Look. *Journal for General Philosophy of Science* 52, 157–170.

Ismael, J. T. (2016) *How Physics Makes us Free*, Oxford University Press.

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.

McCann, H. J. (1998). *The Works of Agency: On Human Action, Will, and Freedom*, Ithaca: Cornell University Press.

Moore, C. (1990). Unpredictability and Undecidability in Dynamical Systems, *Physical Review Letters*, 64(20), 2354-2357.
Murphy, N., Ellis, G. F. R., O'Connor, T. (eds) (2009). *Downward Causation and the Neurobiology of Free Will.* Springer.
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/>.
Pitowsky, I. (1996). Laplace's demon consults an oracle: The computational complexity of prediction. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 27, 161–180.
Popper, K. and Eccles, J. (1977). *The Self and its Brain*. New York: Springer.
Richardson, D. (1968). Some Undecidable Problems Involving Elementary Functions of a Real Variable. *The Journal of Symbolic Logic*, 33(4), 514-520.
Scheffel, J. (2020). On the Solvability of the Mind–Body Problem. *Axiomathes*, 30, 289–312. 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.
Wolfram, S. (1985).Undecidability and intractability in theoretical physics. *Phys. Rev. Lett.* 54(8)735