(preprint, 26 July 2021)

A mechanism that realizes strong emergence

J. H. van Hateren

Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence,
University of Groningen, Groningen, The Netherlands; e-mail: j.h.van.hateren@rug.nl

Abstract

The causal efficacy of a material system is usually thought to be produced by the law-like actions and interactions of its constituents. Here, a specific system is constructed and explained that produces a cause that cannot be understood in this way, but instead has novel and autonomous efficacy. The construction establishes a proof-of-feasibility of strong emergence. The system works by utilizing randomness in a targeted and cyclical way, and by relying on sustained evolution by natural selection. It is not vulnerable to standard arguments against strong emergence, in particular ones that assume that the physical realm is causally closed. Moreover, it does not suffer from epiphenomenalism or causal overdetermination. The system uses only standard material components and processes, and is fully consistent with naturalism. It is discussed whether the emergent cause can still be viewed as 'material' in the way that term is commonly understood.

Keywords Strong emergence · Causation · Materialism · Physicalism · Indeterminacy · Evolution

1 Introduction

Are all causes in nature material in the sense of being ultimately equivalent to law-like interactions of matter-energy, in any combination? The success of the natural sciences over the last centuries suggests that this might be true. The material basis of life is understood in increasing detail, and the same goes for the neural basis of mental processes. Descartes's dualism, which assumed that an immaterial mental substance interacts with matter, has become untenable. Similarly, it is clear that living organisms do not contain a mysterious life force, an élan vital, which was assumed by the vitalists. There is overwhelming evidence, nowadays, that physiological and neural systems are fully composed of physico-chemical processes.

Nevertheless, there are puzzling phenomena in living organisms that seem to have a non-material flavour – phenomena such as agency, intentionality, and consciousness. Are these phenomena indeed completely produced by law-like material interactions? Below it is shown, by construction, that the answer need not be affirmative. The constructed system is maximally simplified in order to make it comprehensible and explainable. It shows that an autonomous cause can emerge from a realizable system that is purely material. The construction depends on a subtle interaction of deterministic and random processes. Moreover, the emergence requires sustained evolution by natural selection. Importantly, the emergent cause is ontological (related to what exists 'out there') and not merely epistemological (related to knowledge and its limitations).

The construction can be fully realized with components at the level of chemistry and basic cellular physiology. Actions and interactions at these levels have been studied extensively, and there is little doubt about the reality of the entities involved (such as molecules and biological cells). The construction thus stays away from lower and higher levels with a more uncertain ontology, such as fundamental physics (where the ontological interpretation of quantum physics is uncertain) and the behavioural and cognitive sciences (where the ontological status of mental phenomena is uncertain). The main point here is to show that well-understood entities can be combined in such a way that an ontologically novel entity emerges that has novel causal efficacy.

The article is organized as follows. Section 2 contains the main result of the article, by constructing and explaining a system that produces a strongly emergent cause. Section 3 discusses the system, in particular with respect to indeterminacy, causation, emergence, and materialism. Section 4 states the conclusion.

2 Construction and explanation

This section presents and explains a minimal material system from which a novel and autonomous causal factor emerges. The system is intended as a proof-of-feasibility: the purpose here is to show that the proposed system works and indeed produces strong emergence. The system is not intended as a faithful model of any existing system. Existing, living systems are far more complex, with more complex heredity and physiology. Moreover, the X process that is assumed in point (2) below has not yet been identified and investigated empirically; hence, its actual existence is not known, even though it is physiologically realizable. Consequently, the system and its evolution have a novel causal structure that has not been studied in biology[1]. The minimal system was proposed and analysed computationally by van Hateren (2015). However, it is analysed here specifically with respect to the question of strong emergence, and it is formulated in a fully non-mathematical, yet self-contained form.

The proposed system combines three causally complex ingredients: evolution by natural selection, random[2] structural change of which the variability is systematically modulated (which is a novel ingredient), and cyclical causation. The statements below are specifically intended to clarify the intricate dynamical feedback structure of the system[3]. They include some redundancy and comments (usually included in sections following ‘Note that’), because a minimal explanation would become incomprehensible if even a single clause were misinterpreted. The explanation of the basic dynamics of the system does not depend on a

[1] It resembles a population of organisms undergoing Darwinian natural selection, with the crucial difference that conventional natural selection depends on a rate of random change (R) that is either fixed or at least not systematically modulated by an internal estimate (x) of each organism’s own evolutionary fitness.

[2] The terms ‘random’, ‘randomness’, ‘determinate’, ‘indeterminate’, ‘nondeterminate’, and ‘deterministic’ are used, throughout this article, in an ontological sense (that is, they do not refer to epistemic uncertainty about the system under consideration, but refer to the presence or absence of intrinsic randomness within the system itself). See also Section 3.1.

[3] Feedback always utilizes implicit time delays, which produces a cyclical rather than a circular causal structure. Therefore, the statements should not be interpreted as a static logical system, because that is likely to produce the buzzer fallacy (Bateson 1979, 58-60, shows that treating the functional description of an electromechanical buzzer as a logical system produces ‘if P, then not P’). In particular, the recursion that the system seems to contain (by utilizing an estimate of fitness to produce a gradual increase of fitness) is only apparent, because of time delays and differences in timescales (see also Section 3).

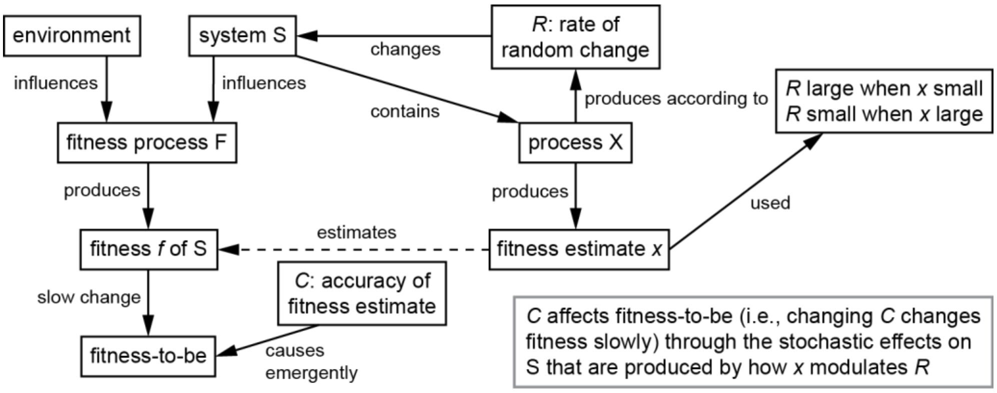

Fig. 1 Overall structure of the proposed mechanism. See the main text for explanations

particular conception of causality, but uses whatever seems clearest. Several points require more discussion than would be consistent with a lucid presentation. Such points are deferred to Section 3. As a guide to the reader, Fig. 1 illustrates the overall structure of the mechanism and its causal consequences. However, a diagram of this kind cannot depict the dynamic and stochastic aspects of the system, and several details are omitted for the sake of clarity. The reader should therefore rely primarily on the explanations below.

(1) Assume a collection of systems S of various forms, embedded in a changing environment. The systems have a fixed lifetime and reproduce continuously. They are being modified continually, gradually, and randomly, each at a variable rate *R*. The fitness *f* of system S is defined as a variable that quantifies its capacity and tendency to reproduce. The fitness of each system varies from moment to moment, depending on system modifications and on environmental change. The latter is assumed to happen continually and to be partly random. Fitness *f* can be thought to be produced by a fitness process F, which properly combines all relevant factors and dynamics within S and its environment.

Note that each individual system varies continually throughout its lifetime, that fitness varies during that time as well (i.e., as a propensity to reproduce at each point in time, and not as a post-mortem reading of actual reproduction), and that fitness is defined here as a variable that applies to an individual system (i.e., not to a population of systems and not to specific traits of a system). It is here not necessary to specify which structural modifications of a system S (compared to the parent of S) are heritable and are thus transferred to the offspring of S. However, at least some such transfer of modifications is needed in order to enable evolution by natural selection.

(2) Assume an internal process X that has gradually evolved within S. X has a time-varying output value *x* that modulates the rate *R* by which S is modified. This is accomplished through conventional causal mechanisms (e.g., by modulating the efficiency of repair mechanisms of random changes of S). Modulation is performed in such a way that *R* is a decreasing function of *x*, that is, *R* is high when *x* is small, and *R* is low when *x* is large.

Note that *R* quantifies the rate of random, undirected modifications, again occurring throughout the lifetime of S. This implies that large *R* produces large variability in the structural changes of S, and small *R* produces small variability. Thus, *x* modulates how much structural variation of S one can expect: much variation when *x* is small and little variation when *x* is large. Because X is part of S, *x* produces structural variation of X as well. Therefore, X varies continually during the lifetime of S.

(3) Systems S drift randomly through the (high-dimensional and abstract) space of possible structural forms (abbreviated to ‘form-space’ below). This drift through form-space occurs continually, because systems are being modified continually. It follows from (2) that systems are more likely to linger at forms that produce a large *x* than at forms that produce a small *x*. This is so, because systems drift away more quickly from forms with small *x* (since the rate of change, *R*, is high there) than from forms with large *x* (since *R* is low there). Purely for statistical reasons, forms that produce large *x* appear sticky, whereas forms with small *x* appear repellent.

Note that the above mechanism produces clustering (in form-space) around forms that produce large *x*, at least when observed at the level of a population of systems. Individual systems only cluster in a probabilistic sense, by having an increased likelihood of acquiring and sustaining a form with large *x*. Because the mechanism is purely statistical, a system only gradually responds to changes of X and *x* (that is, changes in which parts of form-space produce large *x*; such changes can be produced by environmental changes of the input to process X as well as by changes of the structure of X). Thus, the relation between cause (*x*) and effect (probabilistic clustering) is gradual and slow: it takes time.

Formally, drifting through form-space is analogous to how molecules drift—in a random-walk-like manner—through a fluid or gas in a diffusive process (e.g., of a drop of ink in water). Such processes are not instantaneously effected, but slowly and gradually. More specifically, the current mechanism would be analogous to diffusion in a volume in which there is structure in the speed of diffusion, such as a volume of water with zones of different temperatures. Ink particles would then gather mostly in the zones with low temperature, because they are expelled more quickly from the zones with high temperature (which produce a higher speed of diffusion).

(4) Assume further that X has evolved to be, in effect, an estimator that produces *x* as an estimate of the fitness *f* of S. The term ‘estimator’ is used here in the modern statistical sense of a method (here realized as the process X) that produces an estimate (here *x*) of the value of a variable (here *f*). The process X may, for example, get input from evolved sensors about environmental factors relevant for fitness. Such factors are then used to produce an *x* that is likely to have, at least roughly, a similar value as *f*, and that is likely to mimic, at least roughly, how *f* varies from moment to moment. The similarity (or ‘correspondence’ in the colloquial sense) between *x* and *f* is denoted and quantified by *C*, that is, *C* quantifies how well *x* estimates *f*. Another way to state this is that *C* denotes the level-of-fit that *x* has to *f*. Here *x* and *f* usually vary across time because of environmental change, and *C* might gradually change as well because of changes in X or F. If *C* is small, then *x* is a poor estimate of *f*. If *C* is large, then *x* is a good estimate of *f*. If *C* is negligible (zero or close to zero), then *x* and *f* are unrelated.

Note that X and what it does are introduced here and in (2) by assumption. However, in (7) below it is argued that X and what it does are evolvable and sustainable (see also van Hateren 2015). This makes the presence of X in a form similar to what is assumed here consistent with evolution by natural selection, and far from implausible.

An analogy may help to understand how X and *x* are related to F and *f*. One can think of F as the weather (a dynamical process) with *f* then a single time-dependent variable of that process (e.g., the temperature at a particular place). Then X would be a simulation of the weather (such as running in a computer) with *x* the resulting estimate of the temperature at that same particular place. *C* then quantifies how well the simulation estimates that temperature, including how it changes over time. An accurate estimate means a large *C*, that is, it means a high level-of-fit that *x* has to *f*. Both the temperature *f* and its estimate *x* usually change across time. In addition, *C* itself might change across time as well. For example, *C* could increase when the simulation is improved (by refining the computer program), and it might decrease when the climate changes, which might render the assumptions that were used for the simulation less accurate.

(5) Because *x* estimates fitness according to (4), large *x* is correlated with large fitness. A system is more likely, then, to reproduce during the time that it lingers, according to (3), at forms with large *x*. In effect, process (3) combined with (4) biases each system to spend more time at forms with large fitness than at forms with small fitness. As a result, this enhances the time-averaged likelihood that a system reproduces, that is, its average fitness is enhanced. This effect on the fitness of individual systems is gradual and slow, because it depends on the statistics of drifting through form-space. In order to stress the fact that fitness is not immediately affected but gradually, the resulting fitness will be denoted by ‘fitness-to-be’ below. ‘Fitness’ and *f* still refer to current fitness.

Note that reproduction is increased here by combining two quite different mechanisms. The first mechanism, (2), utilizes conventional material causation. It increases the likelihood of having systems with large *x* according to (3). This mechanism would also work if *x* were unrelated to fitness, and it does not influence fitness-to-be by itself (i.e., the gradual increase of fitness is not a direct effect of *x*). The second mechanism, (4), is unconventional. It depends on *x* estimating *f*, which is denoted and quantified by *C*. By itself, this estimate (or its accuracy) has no direct effect on fitness-to-be either (because it lacks, by itself, a mechanism that can affect the structure of S and X). Separately, neither of the two mechanisms affect fitness-to-be. But in combination they do.

(6) The efficiency of mechanism (5) is enhanced when the strength of the similarity *C* between *x* and fitness is increased, such as may result from random modifications of X in a system. The efficiency is enhanced, because a system with increased *C* is even more likely to reproduce when it lingers, according to (3), at forms with large *x* (since large *x* combined with a large *C* implies a higher fitness than large *x* combined with a small *C*). Therefore, mechanism (5) implies that, on average, an increase of *C* leads to an increase of fitness-to-be, that is, a higher fitness-to-be than would occur when *C* would not have been increased.

(7) As already noted in (5), lingering at forms with large *x* would have no effect on reproduction if *x* were unrelated to fitness. Therefore, the similarity *C* between *x* and fitness is required for producing the slow effect on reproduction (i.e., the effect on fitness-to-be). Perturbing *C* tends to change the fitness-to-be that a system will reach, according to (6). Specifically, slightly increasing *C* will slightly increase fitness-to-be, and slightly decreasing *C* will slightly decrease it. Changing fitness-to-be in this way tends to change actual reproduction. Therefore, *C* is a factor with causal efficacy, that is, *C* is a cause. This argument complies with the standard (interventionist) use of causation as is discussed in Section 3.2.

More specifically, one might perturb *C* by changing the structure of the processes X or F, or by changing aspects of the environment that affect *x* (through X) and *f* (through F) in different ways (see further the fourth paragraph of Section 3.2). For example, when the structure of X is changed such that *C* is increased (which is analogous to improving the computer program that simulates the weather), it just means that *x* then better estimates *f* (i.e., the level-of-fit that *x* has to *f* is then higher). It does not, by itself, change *f* (in the same sense as improving a weather simulation will not change the weather). But when *x* better matches *f*, the statistical clustering—towards large *x*—that results (via *R*) from mechanism (3) will better align with fitness; the better aligned clustering subsequently facilitates and enhances fitness. This happens in a slow and gradual way because the clustering is slow, that is, it affects fitness-to-be. Alternatively, when the structure of X is changed such that *C* is decreased, alignment gradually decreases, and as a result fitness does so too.

Note that *C* does not directly affect fitness (nor does it directly affect X, *R*, or S). It affects fitness only indirectly, slowly, and gradually, by enhancing the time-averaged likelihood that a system reproduces, through mechanism (5). This enhancement depends on the fact that mechanism (3) produces a basic tendency towards large *x*, with *x* being similar to fitness *f* because the X process has a structure that produces a large *C*. Mechanism (3) works through the conventional causal mechanism (2). Mechanism (2) does not directly depend on fitness, nor does it directly change fitness. Nevertheless, it is evolvable and sustainable, because it indirectly enhances fitness-to-be through (5). Mechanism (4) is evolvable and sustainable as well: the better *x* estimates fitness (i.e., the larger *C*), the higher the resulting fitness-to-be will become according to mechanism (6). Hence, there is selection pressure on X to produce large *C*. Mechanisms (2) and (4) can coevolve, because they reinforce each other. Moreover, initiating their evolution is facilitated by the fact that they are beneficial even when weakly present.

(8) The major conclusion of (7) is that *C* is a cause. *C* gives system S a causal efficacy that still requires the material constitution of X, but that goes beyond the conventional physical efficacy of X as produced by its constituents (including their activities and interactions). The latter only influence *R*, by conventional material causation. However, an effect on fitness-to-be occurs only when a non-negligible *C* is present as well. *C* acts as a cause, but it is a strongly emergent cause (in the sense of Bedau 1997, see Section 3.3). It produces a causal power that is ontologically novel, that is, the causal power does not follow completely from the micro-physical properties of X (nor from those of X and its environment, see Section 3.4). Novel causal powers are indicative of strong emergence. We conclude that *C* is a strongly emergent cause.

(9) Perturbing *C* tends to slowly and gradually change the fitness of S. In particular, increasing *C* typically increases the fitness-to-be of S according to (7). Thus, increasing *C* is likely to have significant and directed material consequences (because it changes reproduction). In other words, the strongly emergent cause *C* can direct matter. In addition, *C* can be changed by material changes, such as in X. Hence, this provides an example where a strongly emergent cause interacts with conventional material causes.

3 Discussion

Points (8) and (9) above show that the proposed material system produces a strongly emergent cause and that this cause can interact with conventional material ones. It is useful to state the more general reasons why this system can accomplish this. The main reasons are related to a cyclical use of nondeterminism, to evolution by natural selection, and to complementary causes. This is discussed further below, as an informal means to help the reader to comprehend the system explained in detail above. But before doing so, a major point is addressed that may appear confusing at first sight (that is, why *C* can have both an estimative and causal aspect). It concerns the basic question of how one should interpret *C*.

Naively, when the system is first defined and before its dynamics are analysed, *C* is simply a standard level-of-fit that *x* has to *f*. Although it is an objective feature of the world (in the sense of being objectively assignable by any competent observer), it does not participate in the dynamics of the world. Thus, it has not much of an ontic standing. It is not an entity in the sense that it could influence, change or cause things. But because of the mechanism that runs from X via *R* to randomly modifying S (see Fig. 1), *C* acquires a novel causal power (directed at fitness-to-be). It is its only causal power. Having this power makes *C* ontic: it becomes an entity in the above sense (see also the final paragraph of Section 3.2). *C* is then still a level-of-fit (and, technically, it still quantifies the accuracy by which *x* estimates *f*), but it is a non-standard level-of-fit, an ontic one with causal power.

A potential source of confusion is that *C* seems to have a dual role in the analysis. On the one hand, it quantifies how well *x* estimates fitness—it is then simply the level-of-fit that *x* has to *f*. On the other hand, Section 2 claims that *C* is, in addition, a cause, namely a factor that affects fitness-to-be. Wouldn't this amount to a category mistake, by using a single quantity, *C*, for two radically different notions (i.e., quantifying estimation and being a cause)? Moreover, estimation might be viewed as a merely descriptive and epistemic tool, presumably objective and real, but without causal clout in the real world. In contrast, the causal aspect of *C* is claimed to be ontological. Actually, there is no erroneous mix-up of categories here, but this is exactly the kind of categorical uncertainty that is to be expected of strong emergence. Prior conceptions of categories cannot suffice, because an entity with a novel causal power cannot appear just out of thin air. Novel causal power has to be acquired by something that is there already, even if this something is not yet a full-blown entity (such as *C* as a standard level-of-fit). A key point here is that the two aspects of *C*, related to estimation and causation, occur at different timescales. The estimative aspect of *C* occurs at a fast timescale, namely the one at which the components of X operate to produce an *x* that mimics *f*. The causal aspect of *C*—by which it affects fitness-to-be—occurs at a much slower timescale, namely the one at which fitness gradually increases through the statistical clustering mechanism of point (3) in Section 2. The two aspects are simultaneously there, but their dynamics occur at disparate timescales. The two aspects of *C* can thus operate alongside without significantly interfering with each other. The process X can easily let the estimate *x* of fitness follow the slow changes of fitness that depend on the clustering. In effect, estimation and causation are decoupled.

The objection that estimation is only a human epistemic tool does not hold either. As is argued in the final paragraph of point (7) in Section 2, *C* is an evolvable cause. Hence, there is no dependence on humans or human epistemology when *C* evolves and acquires objective causal clout in the world. Nevertheless, one might view the estimative aspect of *C* as part of a primitive epistemology that has evolved within system S itself: the fact that *x* is 'about' *f* may be regarded as a primordial form of 'knowledge' (ascribable to S) about the world. Such a minimal epistemology is evolvable, because *C* has a causal aspect that promotes fitness.

The proposed mechanism is nondeterministic (the term 'nondeterministic' is used here and below to denote 'not completely deterministic'; see also Section 3.1), because it depends in a fundamental way on utilized randomness. When systems of the kind considered here are deterministic, any emergent causation may be classified as weak emergence (Bedau 1997; see Section 3.3). Such a classification also applies to most nondeterministic systems, that is, systems that combine determinism with randomness. Randomness usually does not contribute to novel causal powers; thus, involving randomness is not, by itself, sufficient for producing strong emergence. However, the nondeterministic system explained here is of a rather special kind. It couples deterministic and random factors in a cyclical and sustained way, such that it is impossible to separate them and impossible to neglect the effects of randomness. The randomness is even indispensable, because without the randomness there would be no effect on fitness-to-be and *C* would not be a cause. Importantly, the specific way by which deterministic and random factors are coupled (i.e., by how *x* and *R* are related) tends to increase fitness. In combination with selection pressure, this results in a stable factor with a strongly emergent causal efficacy, as explained in Section 2.

Complementary causes are responsible for the fact that the emergent cause *C* can still affect matter. X has two different causal aspects, an emergent and a conventional material one. The emergent causal aspect of X is produced by the fact that *x* is an estimate of fitness (as denoted and quantified by *C*). The presence of a non-negligible *C* is required for producing the effect on fitness-to-be. But actually producing this effect depends crucially on the material causal aspect of X, which is the fact that X is, in addition, a regular physical process occurring within system S. This physical process modulates the rate of change *R* of the system, through conventional material causation (such as by facilitating or counteracting the effects of random events[4]). Yet, this material causation can only affect fitness because of the emergent causal aspect of X. Exactly the same material causation, with the same X and *x*, would have no effect on fitness-to-be when *x* would not mimic fitness any more (for example, when fitness would have been drastically changed by large random disturbances of the environment). Thus X couples the emergent cause with the material cause. These causal aspects of X are complementary, because they are both indispensable. Neither of them can do the job alone. Therefore, the complementary causes produce neither epiphenomenalism, nor causal overdetermination. The mechanism does not require or produce a dualism of the Descartian type. Nevertheless, one might perhaps view it as letting some sort of dualism emerge from a physical monism.

The proposed system is based on several fundamental assumptions, in particular that there is indeterminacy and that causation can be understood in a particular way. Moreover, it has consequences for several fundamental questions, in particular concerning the existence of strong emergence and the correctness of materialism and physicalism. These topics are discussed below.

3.1 Indeterminacy

A major aim of science is to uncover, systematize, and explain the law-like regularities that one can find in reality. However, it is a valid question whether reality is fully produced—according to such regularities—in a determinate way (i.e., whether it is fully deterministic) or that reality is partly indeterminate (such as when some events occur randomly and without sufficient cause). For the construction explained above it is not necessary to consider this question in its most general form, pertaining to all of reality at once. It suffices to consider the constructed system combined with its causally relevant environment (i.e., the environment relevant for producing its evolutionary fitness). Let us call this combination the ‘inclusive system’. Above we assumed that there is some indeterminacy—in addition to law-like regularities—in the constructed and inclusive systems, in the form of random system modifications and (partly) random environmental change. The indeterminacy is called ‘randomness’ here, and it is assumed to be present throughout the systems. Note that ‘randomness’ refers here to temporal indeterminacy, and not to some structural property of a system (in the sense that one might call a spatial pattern less or more random). Although most engineered systems and many biological sub-systems specifically tend to reduce and control randomness (then usually referred to as ‘noise’), the proposed system specifically utilizes it.

It is plausible that, in effect, any constructed or inclusive system indeed contains indeterminacy and is, thus, nondeterministic (i.e., not fully deterministic). This is plausible because no system is completely isolated from the rest of the universe. In classical physics, long-range electromagnetic and gravitational fields that originate from outside the inclusive system inevitably disturb it in an intractable way. Such influences are not negligible (for an extreme example see Berry 1988, who argues that even the unknown whereabouts of a single

[4] Such mechanisms are known in molecular biology; a simple (non-biological) way by which one can understand the concept of modulated randomness is when one would modulate the distance between a radioactive source and a biological organism, and thus modulate the mutation rate in the latter.

electron at the edge of the visible universe produces sufficient gravitational uncertainty to yield intractable molecular motion, here, within a tiny fraction of a second). In quantum physics, systems inevitably couple to their environment and subsequently decohere, producing indeterminate outcomes (as observables, irrespective of how one interprets the ontology of quantum physics). In any case, the constructed and inclusive systems contain indeterminacy that is either strictly ontic or at least indistinguishable from being strictly ontic.

Empirically, indeterminacy is commonly observed in systems. At the submicroscopic scale there are indeterminate quantum events, at the microscopic scale there is thermal noise (random movements and interactions of molecules), and at larger scales there is indeterminacy caused by non-linear and unstable dynamics. The latter can readily amplify (sub-)microscopic indeterminacy to macroscopic scales. Such instabilities are known to be very common in dynamical systems. A dramatic example is a study by Laskar and Gastineau (2009), who show that infinitesimal causes can have runaway effects even in the planetary system. Similarly, it is clear that biological systems (ranging from the cellular to the neural and behavioural level) are nondeterministic to a considerable extent (e.g., Balázsi et al. 2011; Faisal et al. 2008). Indeterminacy is not negligible at any level in such systems.

3.2 Causation

The explanation of the proposed system utilizes cause-and-effect language. Such language has many uses in many different fields, and it has proven difficult to define causation in a way that covers all such uses at once (for an overview of different approaches see Illari and Russo 2014). However, the proposed system belongs to the realm of the basic (matter-oriented) natural sciences, in particular the ones that deal with complex material systems (such as occur, for example, in cellular physiology, neuroscience, and atmospheric science). Within these fields, causality is usually captured by making a mechanistic model of the investigated system. The model—such as the one presented in Section 2—then contains the conjectured causal structure of the mechanisms or processes involved. The importance of mechanisms and mechanistic causation in the sciences has recently become the focus of a range of studies (e.g., Glennan 1996; Machamer et al. 2000; Bechtel and Abrahamsen 2005; Craver 2007; Illari and Williamson 2012; Glennan 2017). Much of this work has a reductive flavour, that is, it is consistent with ontological reduction of causation. However, the term ‘mechanism’ is used in this article in a general sense, not necessarily reductively and not necessarily referring to deterministic systems.

The causal structure of mechanistic systems can be probed either by an empirical intervention (i.e., through an experiment on an actual system), or by a theoretical intervention (i.e., by mathematical or computational analysis of a modelled system). Interventions are typically performed by slightly perturbing (altering) some part of a system. When a system is thus perturbed (or would be perturbed), and subsequently some part of it tends to respond in a statistically reproducible way (or would do so), then one can infer that a causal connection exists. This strategy (perturb and observe the consequences) was used in point (7) of Section 2 to show that *C* is a cause of fitness-to-be. The strategy is standing practice in science, being closely related to interventionist approaches to causation that are discussed in the philosophy of science (Woodward 2013) and in statistics (Pearl 2009).[5]

[5] One might worry that an interventionist approach would be difficult here if there is a causal loop from *C* to fitness and from fitness back to *C*. However, such a loop does not occur here. Fitness does not directly affect *C* (*f* will usually change all the time because of environmental change, but *x* can normally follow that, keeping *C* unaltered; compare this to the fact that the weather normally does not affect the accuracy of a weather simulation). In contrast, *C* affects fitness (but only slowly). The

Yet, there are valuable alternatives to the interventionist approach. One may then wonder whether such alternatives would also classify *C* as a cause. A few examples are briefly discussed here. First, a counter-factual take on causation would indeed classify *C* as a cause: if *C* had been negligible (i.e., if *x* would have been unrelated to *f*), an effect on fitness-to-be would not have obtained either. Again, the effect that is meant here is the gradual and slow effect produced by the drift-related process utilized in mechanism (3) of Section 2. A second possibility is to equate causation with the existence of a productive causal mechanism between cause and effect (see Glennan 2017, Ch. 7, and the remarks above about capturing causation through a mechanistic model). Using this approach again leads to the conclusion that *C* is a cause: there is indeed a (stochastic) mechanism (as explained in Section 2) that produces a change of fitness-to-be when *C* is changed. As a final possibility one may assume that causation requires that something (e.g., energy) is transferred from cause to effect. Both in applied and fundamental physics this is typically the case (for an extended version of the transference theory see Ardourel and Guay 2018). However, it is not immediately clear what *C* might transfer to fitness-to-be. Given the mechanism, whatever is transferred must at least involve something statistical, such as entropy or information. But it would require further analysis to see what is going on in terms of transfer.

The argument in point (7) of Section 2, as well as several of the alternatives discussed above, establish that *C* is a cause. One might object that the true cause may not be *C* but rather the X process (through *x*), because *C* depends on *x*. However, such an argument would ignore the fact that *C* also depends on the process F (through *f*), since *C* denotes how well *x* mimics *f*. A perturbation of *C* could result from perturbing X or F or both. It could even be produced by an environmental change that affects X and F in different ways. The effect of such changes depends on changes in *C*, because mechanism (3) of Section 2 does nothing but letting S cluster around large values of *x*, which can only change fitness-to-be in relation to how similar *x* is to *f* (that is, depending on the value of *C*). No change of fitness-to-be will occur when *C* is not changed. The latter may occur even when X, F, and the environment all change but approximately compensate each other. This is possible because X and F are two separate, physically unrelated processes that can be perturbed independently (compare that to the fact that a weather simulation and the weather can be perturbed independently). The key point here is that any change must always run via a change in *C* in order to affect fitness-to-be. In other words, *C* is the crucial, indispensable causal factor in this mechanism.

The above discussion on causation primarily concerns how scientific knowledge about causes and causation is acquired, that is, it is primarily related to the epistemology of causation. On the assumption of realism, scientific knowledge should somehow correspond to or refer to what actually exists (i.e., to the ontology of reality). This is often unproblematic, such as when referring to well-studied objects and their characteristics (e.g., a specific planet or a specific molecule). But causation and the associated law-like regularities are more problematic, because it is not clear to what extent fundamental physical laws or fundamental causal dispositions are ontological (see, e.g., Mumford and Anjum 2011; Bird 2016). However, at non-fundamental levels (such as used for the system proposed here) this problem can be avoided. When one assumes a mechanistic framework, one can explain what a particular mechanism causes (i.e., which outcome it produces) by the activities and interactions of its components (Machamer et al. 2000; Illari and Williamson 2012; Glennan 2017). Such components and how they (inter-)act are then taken as given facts, with an ontology that need not be known in all its details. This strategy thus avoids entering a long

deeper reason for this asymmetry is that the latter change depends on a driven diffusive mechanism (from *x* via *R* to random changes of S). Such a mechanism is irreversible (for statistical reasons).

regress of explanations at lower and lower levels, of which it is not clear if and how it bottoms out.

If a particular cause can be completely explained by the above strategy, then such a cause can be said to be amenable to ontological reduction in the sense of Van Gulick (2001). For example, it is assumed above that *x*—if interpreted as a cause of *R*—is fully explained by the detailed workings of X. Then the ontology of the cause *x* is reducible to the ontology of the components of X (including their interactions). Similarly, it is assumed that *f*—if interpreted as a cause of reproduction—is fully explained by the ontological details of system S, its environment, and their interactions (i.e., it is explained by the process F). This implies that such causes are not autonomous, but rather are produced by other causes, namely those at the component level. Although such non-autonomous causes are objective and real, they have no independent causal power[6]. The causal power of a non-autonomous cause is then fully defined by the causal powers of its constituents (including their interactions).

However, the ontology of the emergent cause *C*—as affecting fitness-to-be—is different. Its causal power is autonomous in the sense that it is not fully defined by the causal powers of the constituents of the system and its environment. This is so, because an indispensable part of the causal power of *C* is produced by randomness (which has no fixed ontology and does not correspond to constituents with causal powers). The randomness is indispensable, because without it *C* would not be a cause and there would be no effect on fitness-to-be. A cause *C* with autonomous causal power must be an autonomous entity. This is based on the notion that something that has the potential to influence other entities (by having some sort of causal power) has to exist, i.e., has to be an entity. Having causal autonomy (distinctive efficacy) then guarantees ontological autonomy (distinctness), by Leibniz's law (identicals are indiscernible, thus discernibles are not identical); see Wilson (2015, p. 372). In other words, the causal analysis of the proposed system shows that it produces a novel, emergent entity *C* with autonomous causal power.

3.3 Emergence

The mechanism explained above produces emergence. A general discussion of emergence is beyond the scope of this article (but see the anthologies by Bedau and Humphreys 2008 and Gibb et al. 2019, as well as Wilson 2015; Gillett 2016; Humphreys 2016; O'Connor 2020). Nevertheless, it is important to indicate where the proposed mechanism is positioned within the field studying emergence. Guay and Sartenaer (2016) usefully distinguish three dimensions along which one can analyse emergence. First, their ontological-epistemological axis distinguishes ontological emergence (which is 'out there' in the natural world) from epistemological emergence (which is just a consequence of our representation of the natural world). Second, their strong-weak axis corresponds to the extent to which emergence is present fundamentally (my take on what they mean by 'in principle') or only in practice. Finally, their synchronic-diachronic axis corresponds to whether the emergent phenomenon (such as an event, a property, or a process) occurs simultaneously with its constituents (such as events, properties, or processes) or distinctly later in time.

In other literature (e.g., Bedau 1997; Chalmers 2006; Kim 2006), the ontological-epistemological and strong-weak dimensions of Guay and Sartenaer (2016) are often combined into two broad possibilities for emergence, denoted by 'strong emergence' and 'weak emergence' (see also O'Connor 2020). Weakly emergent phenomena are then novel

[6] Having 'causal power' is used in this article in the general and weak sense of having a disposition, tendency, or propensity to produce certain effects; the use of the term here does not assume a specific ontology of powers.

and surprising primarily for epistemological reasons. A system may be so complex that its properties and dynamics cannot be explained in simple terms. In principle, one could simulate the microdynamics of such a system and, thus, reproduce the emergent phenomenon (Bedau 1997, 2008). The phenomenon is then explained computationally, but it is still novel and surprising. Strongly emergent phenomena, on the other hand, are novel and surprising primarily for ontological reasons. In particular, such phenomena produce novel causal powers that are not explainable in terms of those of the components of the system and its environment (Kim 2006, 2010).

The type of emergence produced by the system that is proposed here can be unequivocally positioned on two of the three axes of Guay and Sartenaer (2016): it is ontological and present fundamentally. But the position on the synchronic-diachronic axis is more complex. A purely synchronic relation between constituents and emergent may be viewed as one of dependence or constitution (e.g., Gillett 2016). However, the present study assumes a regular matter-based system, which thus has to be consistent with fundamental physics. All of the fundamental theories of physics contain differential equations of time; hence it is difficult to see how such a system could produce emergence that is both ontological and purely synchronic (see also remarks and references on this topic in Guay and Sartenaer 2016, p. 301). Pure synchronicity depends on connecting properties (or events, or processes) at different positions at the same time, which fundamental physics cannot do (at least not in a sense relevant for the current study).

That the emergence produced by the proposed system is not purely synchronic is also implied by the fact that it focusses on causation (in the sense of producing effects), which—in matter-based systems—inevitably involves time delays. However, it would be inappropriate to characterize the emergence here as diachronic. The constituent causes (involved in how X produces *x* and thus modulates randomness) and the emergent cause (*C*) act continuously and spread out in time. The emergent cause-and-effect relation (between *C* and fitness-to-be) builds up statistically and gradually. It occurs on a much slower timescale than the constituent cause-and-effect relations within X. Yet, these constituents and the emergent cause *C* overlap in time to a considerable degree, that is, they occur almost simultaneously. They do not occur at clearly separable, distinct moments in time (as would be required for pure diachronicity)[7]. Therefore, the emergence is best characterized as near-synchronic. In order to keep the formulations in this article simple, the term 'strong emergence' is used here to refer to the near-synchronic, ontological, and fundamentally present emergence of a causal power. In the terminology of Van Gulick (2001), strong emergence as used here corresponds to a radical-kind emergence of a causal power.

A specific form of emergence, where an ontological transformation occurs across time, is called transformational emergence (Humphreys 2016, Ch. 2; Guay and Sartenaer 2016). One may wonder whether the proposed system could be understood in those terms, as one might think that *C* is transformed from a simple level-of-fit to a cause. However, nothing is transformed here, because *C* remains a level-of-fit and the causal power is just an emergent addition. Moreover, transformational emergence is diachronic, whereas the emergence proposed here is not: *C* is simultaneously a level-of-fit and a cause, even as the dynamics of these two aspects occur at mostly different timescales.

Finally, one may wonder what would be the emergence base from which *C* emerges. From Section 2 and Fig. 1 it is clear that many different parts of the world are participating in

[7] Note that the notion of causation taken here is more general than the traditional notion of event causation (i.e., one distinct event causing another distinct event). This is necessary, because mechanisms (such as the one proposed here) often do not operate with discrete events, but rather through continuous and overlapping influences.

producing *C*: system S and its components, as well as the environment, which coproduces fitness. But a crucial part of the emergence base that is not explicitly represented in Fig. 1 is ontic randomness, without which *C* would not be a cause. Yet, ontic randomness is not definable in a definite way, because its concurrent outcomes are not determined before they occur. Only with hindsight one could reconstruct the randomness and its consequences. But hindsight is a form of epistemology that is not available to the mechanism itself, including during the time that *C* gradually causes fitness-to-be (see also below). Therefore, the concept of a definite emergence base cannot be applied to the proposed mechanism, at least not in an exact way.

3.4 Materialism and physicalism

According to point (8) in Section 2, *C* is a strongly emergent cause. But is *C* a material cause? The answer to this question depends on how one interprets the meaning of the term 'material'. As indicated by the first sentence of the Introduction, we will take it as referring to being fully produced by law-like interactions of matter-energy, in any combination. Thus, it is not restricted to mere matter, but also includes, for example, forces and fields. This is consistent with how the term 'material' is commonly understood in the natural sciences (including by those working on complex material systems). The common interpretation is that a material cause is a cause that would be completely specified by its material constitution, that is, by its material constituents and their activities and interactions[8]. Moreover, the material constitution must be well-defined and must be knowable, at least in principle. This implies that such a cause is amenable to ontological reduction (in the sense of Van Gulick 2001), at least in theory: if one would reassemble the full material constitution from scratch[9], then one would get exactly one's material cause and its efficacy, no less and no more. A second common interpretation of material causation is based on the notion (e.g., Wilson 2015, p. 351) that causation is first of all a cause-and-effect relation between spatiotemporally located goings-on (with 'goings-on' standing for events or processes). Then material causation can be interpreted as a cause-and-effect relation between the material events or processes to which such goings-on correspond (i.e., to which they can be reduced ontologically). If either one of these two common interpretations of material causation does not apply, then it seems best not to regard the causation as material in the conventional sense of the term. As is argued below, neither of these interpretations does in fact apply to *C*.

The first interpretation does not apply because there is no suitable basis for an ontological reduction of *C* (or, in alternative formulations, a basis on which *C* supervenes or a basis by which *C* is realized). One might think that *C* could be reduced to the material process (X) that is required for producing *C* and its effect on fitness-to-be. This might be so, because the value of *x* is fully produced by the material constituents of X, and the effect of *C* on fitness-to-be depends on how *x* modulates *R*, the rate of random change. However, such a reduction cannot work, because *C* depends not only on *x*, but also on fitness *f* (recall that *C* quantifies how well *x* estimates *f*). Fitness depends on interactions of the system S and its environment, which both extend beyond X. Therefore, X is not sufficient for specifying *C*, nor for specifying the effect of *C* on fitness-to-be.

[8] In the context of physicalism, Hüttemann and Papineau (2005) argue for distinguishing part-whole reduction and inter-level reduction (e.g., when connecting the mental and physical levels). No such distinction would apply to the current study, because it focusses on a single, well-understood level (chemistry-based physiology, see Section 1).

[9] For open or nondeterministic systems this could include external contingencies (i.e., randomness) assumed to be independent of the form and dynamics of the system (see Bedau 1997).

Alternatively, one might think that *C* could be reduced to a broadened basis, which includes not only X and S, but also any part of the environment that is relevant for fitness. Such a reduction cannot work either, as is argued next. The effect of *C* on fitness-to-be is stochastic, being produced slowly and gradually by modulated random change. This randomness is assumed to be ontic, which means that each individual change is indeterminate up until the short time interval during which it is produced (this remains true also when the statistics of changes are modulated over time). The specific outcome of each change only becomes determinate during this production, but before that it is fully indeterminate: not just epistemologically (due to lack of knowledge about the system) but ontologically (irrespective of how completely the production could be characterized; see also Section 3.1). The effect of *C* on fitness-to-be is slow, thus it depends on the specific outcomes of a series of individual changes during a stretch of time. The system compounds this effect over time in such a way that the probability distribution of each subsequent change, as well as its micro-effects, depend on the results of previous changes (because these results partly determine how the system subsequently modulates randomness). This means that ontological reduction as defined above is not possible: one could not reassemble the material constitution of the broadened basis—over the stretch of time where *C* assembles its effect on fitness-to-be—from scratch, because randomness prevents reproducing this constitution. This is so, because randomness would produce a different material constitution each time when one would attempt such a reassembly: the constituents and their activities and interactions would not be the same. Consequently, not only the sequence of forms of X and S would be significantly different each time, but also *C* and its effect on fitness-to-be. Therefore, the causal efficacy of *C* is not fully given by its ontological basis, because that basis stretches across time and is partly indeterminate[10]. Of course, one could observe the specific realizations of the random changes of X and S over a stretch of time (after the fact) and use those for faithfully reproducing the cause *C* and its effect, from scratch. But that would be cheating, because a reproduced series of random changes is not indeterminate, but determinate (and thus not random; recall that the terms random, indeterminate, and determinate are all used here in an ontological sense). That would fundamentally change the ontology of the system that is presented here, in effect taking it to be fully deterministic. We must conclude, therefore, that *C* cannot be reduced ontologically, not even in a broadened basis.

The second commonly used interpretation of material causation, that it is a relation between spatiotemporally located goings-on of a material kind, does not apply either. The effect, an increase of fitness-to-be, is a going-on of the right kind, because fitness characterizes a material process (reproduction) that is fully produced—in a complex way—by the physical properties of system S and those of its environment. However, the cause of this increase, *C*, is not a material going-on, as is argued next. *C* denotes and quantifies how well *x* estimates fitness, and this estimation is in fact the crucial condition for *C* acting as a cause. An estimate is, roughly speaking, a relation of some kind[11]. The process that produces *x*, X, is a material going-on, and the same goes for the process that results in *f*. However, the fact that *C* is a cause (by slowly affecting fitness-to-be) is fully attributable to the relation between *x* and

[10] Note that this is not an epistemological issue, but an ontological one: even the system itself could not construct the causal aspect of *C* before that aspect is slowly and gradually produced through a series of changes, because each change remains indeterminate even to the system itself until the short time interval in which it is produced.

[11] In a proper (two-sided) relation, A being related to B entails that B is related to A (though often in a different way). In contrast, an estimate is one-sided: the estimate points to the estimated, but the estimated does not point to the estimate. Here *x* points to *f*, but *f* does not point to *x* (only the former is causally relevant to S, which contains and utilizes *x*, whereas the latter would be causally irrelevant to S).

f as such (see the fourth paragraph of Section 3.2). This relation, as such, is not a material, spatiotemporally located going-on. Essentially, it is a similarity between two values as they evolve over time, somewhat like a correlation (but more constrained, because the specific value of *x* needs to be close to the specific value of *f*, on average). A relation between two values does not consist of material events or processes. There is no way in which a relation between two values is identical to—in the sense of being indistinguishably replaceable by—a material event or process. Relations between values belong to a different category than material events and processes. We conclude that causation here is not a relation between two material goings-on, but is a relation between a relation of some kind (the estimative aspect of *C*) and a material going-on (the material process characterized by fitness-to-be). Therefore, the causal aspect of *C* fails to comply with the second interpretation of material causation.

Because *C* does not comply with both common interpretations of material causation, we must conclude that the cause *C* is best regarded as non-material[12]. Hence, if the proposed system were constructed (or variants of it actually exist), it would be a counter-example against materialism. However, one should not conclude from this that the cause *C* is non-physical as well, as this depends on how broadly one defines the term ‘physical’. One might argue that anything real that exists must be physical, by definition (Strawson 2008), that anything that occurs in the physical realm (including uncaused random events) is physical, and that anything that emerges from that is automatically physical as well. Then *C* would be a physical cause, along with any other possible cause (if one assumes naturalism). Thus, the current proposal is consistent with physicalism, if that is broadly defined.

Nevertheless, the mechanism as proposed here is incompatible with common, more narrow conceptions of physicalism. Arguments against the existence of strong emergence often depend on the thesis that the physical realm is causally closed, such as “Every physical effect has an immediate sufficient physical cause, in so far as it has a sufficient physical cause at all” (Papineau 2009). The ‘immediate’ is included in order to exclude causal chains running indirectly through a non-physical cause, and can be ignored here. The thesis appears to be correct for deterministic processes, and it is silent about uncaused random events (because of the ‘in so far .. at all’). But it is not obvious that the thesis, or any variant of it, is always true of processes that are both deterministic and random. The mechanism explained above is in fact a realizable counterexample against physical causal closure[13]. It clearly has a physical effect, on fitness-to-be and thus on reproduction. It produces this effect by systematically utilizing randomness (via *x* and *R*) to produce deterministic effects (via the structure of S and X) that subsequently modulate further randomness (via *x* and *R*), and so on. This continual cycle of mixing determinism and randomness produces an effect on fitness-to-be that is of mixed origin: both caused and uncaused. The cycle is complex, nonlinear, and nonstationary, making it impossible—not only in practice, but fundamentally—to disentangle the causal effect that *C* has on fitness-to-be in terms of physical causes and uncaused randomness. More specifically, the cause *C* produces its effect on fitness-to-be by an inseparable composite of physical causes and uncaused randomness. One might think that the ‘in so far as it has a sufficient physical cause at all’ could deal with this, but that is not so. As the definition stands, ‘in so far .. at al’ seems to be equivalent to ‘if’, and is intended to exclude ontic randomness (Papineau 2009). It does not deal with causation of mixed origin. We might change the definition by removing the ‘at all’, and hope that ‘in so far as’ (i.e., ‘to the extent that’) works for the current mechanism. However, this would tacitly assume that one can

[12] Material causation is undermined here even if one thinks that only one of the two interpretations is convincingly discredited.

[13] At least for any physical system that is not equal to the entire universe at once; the latter is not sufficiently well understood to presume that it can be regarded as a regular closed system.

separate the effect on fitness-to-be in a caused and an uncaused part, which is not the case. Thus, ‘in so far as it has a sufficient physical cause’ is inapplicable here; it is meaningless with respect to the proposed mechanism.

Still, some form of physical closure might be formulated by saying something like “Every physical going-on that is caused, is caused by nothing other than physical goings-on or causes that are strongly emergent from physical goings-on.” The term ‘physical going-on’ denotes here any form of physical change, irrespective of whether it is determinate, random, or mixed. Then the physical domain is still closed in some sense, but it is not necessarily causally closed (because any strongly emergent cause, such as *C*, would be a metaphysical novelty, though still physical if that is broadly defined).

Van Gulick (2001) states that mainstream (‘atomistic’) physicalism assumes two core principles, namely AP1 “the features of macro items are determined by the features of their micro parts plus their mode of combination” and AP2 “The only law-like regularities needed for the determination of macro features by micro features are those that govern the interactions of those micro features in all contexts, systemic or otherwise.” The mechanism presented here conforms with AP1, but not with AP2. For example, the process X is, as a matter-based process, fully defined by how its micro items are combined (as in AP1). However, the interactions of its micro features are not the only law-like regularities that are needed for understanding the causal efficacy of X (contra AP2). In addition, the emergent law-like regularity “the factor *C*, which depends on X, affects fitness-to-be” is needed. This regularity is not fully defined by X, nor by X and its context (i.e., by the inclusive system). It is a slow regularity that is established stochastically and gradually, and that depends on how well *x* keeps estimating *f* over time. Thus, it depends on how the inclusive system and the environmental statistics change over time, potentially a long time into the future. Of course, S cannot really tell the future, but has to rely here on the predictive qualities of its X process, the structure of which was gradually established—through natural selection—over time, potentially a long time into the past. This deviates from the standard assumptions of mainstream physicalism (as well as from those of a non-mainstream version as in Papineau 2008). Such standard assumptions are that the causal fate of any system is fully determined by the current system and its current environment, plus possibly some current external contingencies (for open or nondeterministic systems; see Bedau 1997). An explicit and irreducible statistical dependence on the (nondeterminate) future form of a system and the (nondeterminate) future environmental statistics, as applies to the mechanism explained here, is not part of the standard view. Nevertheless, the proposed mechanism is fully consistent with standard science.

4 Conclusion

The mechanism constructed above shows that strong emergence—taken here to be the near-synchronic, ontological, and fundamentally present emergence of a causal power—is feasible in a universe that appears to be fully based on micro-physical laws. This is enabled by the fact that randomness escapes such laws, at least in its detailed realizations (i.e., its actual outcomes). The proposed mechanism takes advantage of this nondeterministic loophole in the apparent law-like nature of reality. An internal estimator of a system’s own reproductive fitness thus produces a slow and gradual increase of fitness, by stochastic means. As a result, the estimate of fitness obtains a novel quality, namely that of being a—strongly emergent—cause of fitness-to-be. It thus provides an example of how strong causal emergence can be realized in a material system. If it actually exists in one form or another, it may explain several of the more puzzling phenomena that occur in living organisms (see van Hateren 2019, 2021 for specific elaborations).

Acknowledgements I would like to thank the reviewers for their constructive criticism and helpful suggestions.

References

Ardourel, V., & Guay, A. (2018). Why is the transference theory of causation insufficient? The challenge of the Aharonov-Bohm effect. *Studies in History and Philosophy of Modern Physics, 63,* 12–23.

Balázsi, G., van Oudenaarden, A., & Collins, J. J. (2011). Cellular decision making and biological noise: From microbes to mammals. *Cell, 144,* 910–925.

Bateson, G. (1979). *Mind and nature: A necessary unit*. New York: E. P. Dutton.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanistic alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36,* 421–441.

Bedau, M. A. (1997). Weak emergence. In J. Tomberlin (Ed.), *Philosophical perspectives: Mind, causation, and world, vol. 11* (pp. 375–399). Malden, MA: Blackwell.

Bedau, M. A. (2008). Is weak emergence just in the mind? *Mind & Machines, 18,* 443–459.

Bedau, M. A., & Humphreys, P. (2008). *Emergence: Contemporary readings in philosophy and science*. Cambridge: MIT Press.

Berry, M. V. (1988). The electron at the end of the universe. In L. Wolpert & A. Richards (Eds.), *A passion for science* (pp. 39–51). Oxford: Oxford University Press.

Bird, A. (2016). Overpowering: How the powers ontology has overreached itself. *Mind, 125,* 341–383.

Chalmers, D. (2006). Strong and weak emergence. In P. Clayton & P. Davies (Eds.), *The re-emergence of emergence: The emergentist hypothesis from science to religion* (pp. 244–254). Oxford: Oxford University Press.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience, 9,* 292–303.

Gibb, S., Hendry, R. F., & Lancaster, T. (Eds.). (2019). *The Routledge handbook of emergence*. London: Routledge.

Gillett, C. (2016). *Reduction and emergence in science and philosophy*. Cambridge: Cambridge University Press.

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44,* 49–71.

Glennan, S. (2017). *The new mechanical philosophy*. Oxford: Oxford University Press.

Guay, A., & Sartenaer, O. (2016). A new look at emergence. Or when *after* is different. *European Journal for Philosophy of Science, 6,* 297–322.

Humphreys, P. (2016). *Emergence: A philosophical account*. Oxford: Oxford University Press.

Hüttemann, A., & Papineau, D. (2005). Physicalism decomposed. *Analysis, 65,* 33–39.

Illari, P., & Russo, F. (2014). *Causality: Philosophical theory meets scientific practice*. Oxford: Oxford University Press.

Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms *across* the sciences. *European Journal for Philosophy of Science, 2,* 119–135.

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

Kim, J. (2010). *Essays in the metaphysics of mind*. Oxford: Oxford University Press.

Laskar, J., & Gastineau, M. (2009). Existence of collisional trajectories of Mercury, Mars and Venus with the Earth. *Nature, 459,* 817–819.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67,* 1–25.

Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. Oxford: Oxford University Press.

O'Connor, T. (2020). Emergent properties. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2020 Edition). https://plato.stanford.edu/archives/fall2020/entries/properties-emergent/.

Papineau, D. (2008). Must a physicalist be a microphysicalist? In J. Hohwy & J. Kallustrup (Eds.), *Being reduced: New essays on reduction, explanation, and causation* (pp. 126–148). Oxford: Oxford University Press.

Papineau, D. (2009). The causal closure of the physical and naturalism. In A. Beckermann, B. P. McLaughlin, & S. Walter (Eds.), *The Oxford handbook of philosophy of mind* (pp. 53–65). Oxford: Oxford University Press.

Pearl, J. (2009). *Causality: Models, reasoning, and inference (2nd edition)*. Cambridge: Cambridge University Press.

Strawson, G. (2008). *Real materialism and other essays*. Oxford: Clarendon Press.

Van Gulick, R. (2001). Reduction, emergence and other recent options on the mind/body problem: A philosophic overview. *Journal of Consciousness Studies, 8,* 1–34.

van Hateren, J. H. (2015). Active causation and the origin of meaning. *Biological Cybernetics, 109,* 33–46.

van Hateren, J. H. (2019). A theory of consciousness: Computation, algorithm, and neurobiological realization. *Biological Cybernetics, 113*, 357–372.

van Hateren, J. H. (2021). Constructing a naturalistic theory of intentionality. *Philosophia, 49,* 473–493.

Wilson, J. (2015). Metaphysical emergence: Weak and strong. In T. Bigaj & C. Wüthrich (Eds.), *Metaphysics in contemporary physics* (pp. 345–402). Leiden: Brill.

Woodward, J. (2013). Causation and manipulability. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019 Edition). https://plato.stanford.edu/archives/sum2019/entries/causation-mani/.