



# Irrational methods suggest indecomposability and emergence

Hamed Tabatabaei Ghomi<sup>1</sup>

Received: 20 June 2022 / Accepted: 24 November 2022  
© The Author(s) 2022

## Abstract

This paper offers a practical argument for metaphysical emergence. The main message is that the growing reliance on so-called irrational scientific methods provides evidence that objects of science are indecomposable and as such, are better described by metaphysical emergence as opposed to the prevalent reductionistic metaphysics. I show that a potential counterargument that science will eventually reduce everything to physics has little weight given where science is heading with its current methodological trend. I substantiate my arguments by detailed examples from biological engineering, but the conclusions are extendable beyond that discipline.

**Keywords** Emergence · Irrational methods · Biological engineering · Synthetic biology · Metaphysical emergence · Indecomposability

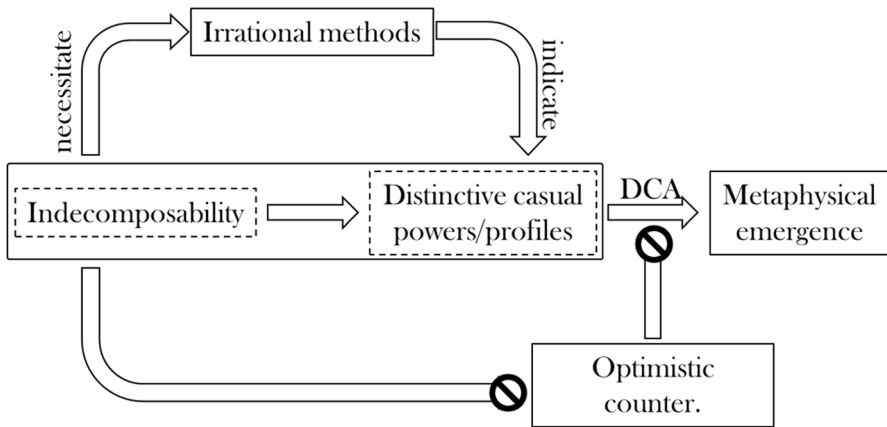
## 1 Introduction

Atoms make molecules, molecules make chemical systems, chemical systems make biological systems, and so it goes step by step all the way up to psychological and social systems (Oppenheim & Putnam, 1958). This reductionistic layered metaphysics that describes everything as ultimately nothing but the pushing and pulling of atomic stuff is the working metaphysics of science (Humphreys, 2016a). In practice, however, science has had limited success in reducing the higher-level scientific models, concepts, and causal relations to those of lower fundamental levels (Fodor, 1974, 1997; Mitchell, 2009; Kaiser, 2017; Mazzocchi, 2008). Even within physics, reduction of higher-level physical phenomena to the most fundamental level has faltered in some cases (Batterman, 2001, 2005). Supporters of the reductionistic metaphysics have to somehow explain away the prevalent non-reductionistic nature of modern science, and one important strategy in their repertoire is taking recourse to what I

---

✉ Hamed Tabatabaei Ghomi  
ht396@cam.ac.uk

<sup>1</sup> Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge CB2 3RH, UK



**Fig. 1** Overview of the arguments

call the *optimistic counterargument* (Barwich, 2021; Bickle, 2006, 2020; Hempel & Oppenheim, 1948). Roughly, the counterargument suggests that although science has so far failed at reducing everything to fundamental physics, it will eventually succeed, or at least it is highly probable that it will, or it is in principle possible. This paper aims to weaken this counterargument and provide support for alternative non-reductionist metaphysical views grouped under *metaphysical emergence* (Wilson, 2021). Along the way, the paper links some scientific methodological choices to the non-reductionistic metaphysics of objects of science.

My approach follows the idea promoted by philosophers such as Cartwright (2007) and Mitchell (2012) that our metaphysical assertions and our views about the future of science could, and should, be based on both historical and contemporary facts about the practice and theory of well-developed science. I argue that, contrary to the optimistic counterargument, the ongoing trend of science shows that it is becoming more and more holistic, and this trend is better explained by non-reductionistic metaphysical views such as metaphysical emergence (Humphreys, 2016b; Wilson, 2015, 2021).

My argument is based on the fact that some domains of science rely on so-called irrational methods. Irrational methods are those that rely on trial-and-error without reference to a clear mechanistic model, and are opposed to rational methods that are based on some theory about the underlying mechanisms in the system under study. Suppose that we observe that science is relying more and more on irrational methods. I argue that over time, this trend warrants the belief in metaphysical emergence as opposed to the reductionistic metaphysics. The penultimate part of the paper (Section 6) provides more details.

I start, in Section 2, by explicating the notion of metaphysical emergence. This sets the theoretical grounds for my proposal that the ongoing trend of science warrants one's belief in metaphysical emergence. The backbone of the following arguments in Sections 3 to 6 is schematically shown in Fig. 1. In Section 3, I discuss what I call the *distinctive causal powers argument* (DCA), which is a line of argument

commonly cited in support of metaphysical emergence. In Section 4, I explain how DCA is blocked by the optimistic counterargument and discuss different versions of this blockade. In Section 5, I first show how so-called *indecomposable* systems make good cases for DCA and weaken the optimistic counterargument. I then go to some real-world examples and argue that the heavy reliance on irrational methods in biological engineering indicates that biological systems are inherently indecomposable. Finally, in Section 6, I generalise and argue that a constant trend of heavy reliance on irrational methods warrants metaphysical emergence.

In general, I propose what I see as a practical argument for metaphysical emergence, in the spirit of Hacking's well-known practical argument for realism. Hacking (1983) famously argued, "If you can spray them, then they're real". I argue that if you have to go irrational, the system is probably metaphysically emergent. In short, I propose: *Irrationality suggests indecomposability. Indecomposability implies emergence.*

## 2 Metaphysical emergence as rejection of generative atomism

The term *emergence* means too many different things in the literature, as are its varieties such as epistemological, and metaphysical.<sup>1</sup> It is, therefore, important to clarify what we mean by the term and its varieties in the context of the present discussion. I use Humphreys' (2016a) definition of emergence as the starting point. Humphreys defines emergence as any sort of violation of *generative atomism*. Generative atomism is the assumption that everything in the world can be reduced to the spatiotemporal arrangements of some fundamental entities and their properties. The fundamental entities of generative atomism are called "atoms," although they might not correspond to what we recognize as chemical atoms. Atoms here simply refer to the most fundamental physical entities, whatever they are. Atoms are both type and token distinguishable, meaning that different kinds of atoms and different instances of those kinds can be identified and individuated. Also, the essential and the non-relational properties of atoms are immutable.

Generative atomism describes the relation between atoms and everything else in two ways, synthetically and analytically. Synthetically, or bottom-up, the collection of atoms and their set of fixed fundamental causal powers are the sole constituents of all other non-atomic entities and their causal powers. And analytically, or top-down, any non-atomic entity can be uniquely decomposed to its constituting atoms with some fixed decomposition scheme. All in all, we can understand generative atomism in terms of a child's Lego game in which everything is simply an assemblage of a fixed and limited variety of Lego pieces following some rules of assembly. The metaphysical content of such a world, so generative atomism goes, consists of nothing

---

<sup>1</sup> For instance, compare the concepts of weak and strong emergence in (Chalmers, 2008; Wilson, 2015; Tabatabaei Ghomi, 2022). Most importantly, one must note that weak and strong emergence usually correspond to epistemic and metaphysical emergence respectively, while weak and strong *metaphysical* emergence are two varieties of metaphysical emergence.

but Lego pieces. Any apparent construction beyond Lego pieces is simply a figment of the child's imagination.

Tied to physicalism, the claim that atoms are necessarily of physical nature, generative atomism forms the working metaphysical assumption underlying modern science (Humphreys, 2016a; Oppenheim & Putnam, 1958). Because physicalism and generative atomism are so intertwined in the scientific mindset, it is important to emphasize their independence. As we will see below, there are accounts of emergence that endorse physicalism and yet reject generative atomism. It is also possible to reject physicalism but accept generative atomism as it seems to be the case in some varieties of panpsychism (Nagel, 2012). When arguing against generative atomism and hence, for emergence, one is not necessarily arguing against physicalism. Emergence is violation of generative atomism and as such, it is in principle compatible with physicalism.

There are two main branches of emergence, metaphysical and epistemological. Metaphysical emergence encompasses all the views that reject generative atomism as a metaphysical fact of our world. Epistemic emergence, on the other hand, encompasses all the views that accept generative atomism as a metaphysical fact, but suggest that this fact slips our epistemological grasp either temporarily (Hempel & Oppenheim, 1948), or permanently (Bedau, 1997; Huneman, 2008).

Various forms of epistemic emergence simply refer, one way or another, to the fact that we, the children who play in this Lego world, cannot understand or describe Lego artifacts in terms of Lego pieces. If we could, we would see that the artifacts we were recognizing as individual "things" were in fact nothing above and beyond their constituent Lego pieces. As the metaphysics of the Lego world exists independently of our epistemic views about it, epistemic emergence is perfectly compatible with generative atomism as a metaphysical fact. Therefore, if we strictly stick to our definition of emergence as violation of generative atomism, it might be better to take epistemic emergence as a variety of anti-emergentism.

Unlike epistemic emergence, however, all sorts of metaphysical emergence clash with generative atomism, even though the nature of this clash is different across the varieties of metaphysical emergence (Alexander, 1920; Batterman, 2001; Chalmers, 1996, 2008; Humphreys, 1997, 2016b; O'Connor & Wong, 2005; Van Cleve, 1990; Wilson, 2021). With the exception of epi-phenomenalist views of emergence such as (Chalmers, 1996), the varieties of metaphysical emergence are unanimous in associating some sort of causal uniqueness with higher-level non-fundamental entities with respect to their fundamental bases. Details of the varieties of metaphysical emergence and the nuances of how they differ are irrelevant to our current discussion as I believe the arguments of this paper can be invoked to support any of these views against epistemological emergence or other sorts of anti-emergentism. But it is important to at least distinguish two main sub-types, namely weak and strong, so that it becomes clearer what we argue for when we argue for metaphysical emergence.

Accounts of metaphysical emergence fall into two general sub-types, weak and strong (Wilson, 2015, 2021). Though in different ways, both types attribute distinctive causal characters to the emergent entities and recognize them as metaphysically different from their lower-level bases. The key difference between the two sub-types,

however, is that weak metaphysical emergence endorses physicalism, but strong metaphysical emergence does not. According to strong metaphysical emergence, emergent entities are of non-physical nature and show non-physical causal powers. The most prominent examples are mind (O'Connor & Wong, 2005), life (Alexander, 1920), and free will (Wilson, 2021). According to weak metaphysical emergence, on the other hand, fundamental physical causes and entities are the only building blocks of our world, but despite this fact, emergent phenomena have distinct causal and metaphysical characters that are different from their constituent building blocks (Wilson, 2015, 2021). The most commonly cited examples of weak metaphysical emergence are objects of special sciences such as biology and chemistry.

In the next section, I discuss how one can argue for the existence of weak or strong emergence by reference to the allegedly distinctive causal powers of higher-level phenomena via what I call *the distinctive causal powers argument* (DCA). The difference between weak and strong metaphysical emergence will become clearer along that discussion.

### 3 The distinctive causal powers argument

One common way to argue for the existence of metaphysical emergence is by reference to some sort of causal uniqueness on the emergent level, compared to the lower, more fundamental levels. This uniqueness can be in the form of novel non-physical causal powers associated with strong metaphysical emergence (Humphreys, 1997, 2016b; O'Connor & Wong, 2005), or distinctive causal profiles associated with weak metaphysical emergence (Wilson, 2015). Both types of causal uniqueness are supposed to be incompatible with generative atomism. The clash between the novel causes of strong metaphysical emergence and generative atomism is obvious. Higher level non-physical causal powers doubly violate generative atomism. First, they show that the causal powers of atoms are not the exclusive governing rules of our world. Second, by implying the existence of some emergent entities that possess and instantiate non-physical causal powers, these causal powers show that atomic entities do not exhaust the metaphysics of our world.

Similarly, the distinctive emergent causal profiles of weak metaphysical emergence also violate generative atomism. An example of a distinct causal profile is where the emergent phenomenon has only a subset of the causal powers that its lower-level fundamental base has. Suppose that phenomenon E is generated by fundamental base B. Were E nothing but B, then the causal powers shown by E (i.e., its causal profile) should be identical to the causal powers of B. But according to weak emergentists, if E is emergent, it shows a distinctive causal profile that is constantly and reproducibly different from that of B. This means that one can distinguish E from B by reference to its distinctive causal profiles. According to Leibniz's Law, identicals are indiscernible (Forrest, 2020) and therefore by modus tollens, if one can distinguish E from B, one can conclude that E is not identical to B (Wilson, 2021). Generative atomism, therefore, does not hold because the metaphysics of the world is not exhausted by only B, but also contains E.

Weak metaphysical emergentists claim that many phenomena of the special sciences are instances of such Es. For example, consider the biological structure and function of a protein (E) in comparison to its amino acid sequence (B). Proteins are polymers of amino acids that fold into specific three-dimensional structures and perform specific biological functions. The function of a protein is primarily determined by its amino acid sequence. Yet, we observe that proteins with markedly different amino acid sequences all fold into similar structures and show similar functions. This is one of the reasons that, from the biological standpoint, proteins are recognised not just by their sequences, but also by their functional and structural characters. Biologists classify proteins into families with markedly similar 3D structures and functional roles, where the sequence similarity between closely related proteins within a family can be as low as 40%, and can be even lower among proteins of large superfamilies (Orengo & Thornton, 2005).

In Horgan's (1989) terms, proteins have specific *quausal* profiles, causal effects *qua* being certain things, which are distinguishable from their general *causal* effects. A weak metaphysical emergentist summons Leibniz's Law here and concludes that members of a protein family as instances of Es are metaphysically distinct from the Bs, which are the amino acid sequences that happen to be grouped in that family. Recognising the reality of these new metaphysical entities fits the significant explanatory role that protein families play in understanding the natural origins and relations of proteins.

In summary, both strong and weak metaphysical accounts of emergence associate higher-level phenomena with distinct causal profiles and use that to argue for the existence of metaphysical emergence.<sup>2</sup> This general line of argument (DCA) starts from the claim of distinctive causal powers or profiles for higher-level phenomena and concludes in claiming the existence of metaphysical emergence. DCA is flexible with respect to one's preferred philosophical understanding of causation. The nature of distinctive causal powers or profiles needed for DCA is merely nomologically motivated and philosophically lightweight. As Wilson (2015) puts it, talk of causal powers here simply refers to the sense that "a magnet attracts nearby pins in virtue of being magnetic, not massy; a magnet falls to the ground when dropped in virtue of being massy, not magnetic." (354) Thus, DCA works almost regardless of one's position on the nature of causation. Even Humeans can construct their own version of DCA.

DCA is particularly suitable if one aims to approach the metaphysical question of emergence from a scientific perspective. Science is most useful in identifying causal relations. So, a good metaphysical argument inspired by science would be an indirect one via a discussion of causal relations. DCA is such an argument. The argument, however, loses its power in the face of the optimistic counterargument, i.e. the claim that higher-level causal powers are merely epistemic artefacts of our limited

---

<sup>2</sup> These causal powers have been so critical for metaphysical emergence that denying their possibility has become the most important line of argument against the existence of metaphysical emergence (Kim, 1998, 1999).

understanding of the fundamentals. This is where the arguments of this paper come to the aid of the emergentist.

#### 4 The optimistic counterargument

Generally speaking, science is replete with systems with higher-level causation that seem to be distinct from, and irreducible to, the fundamental causes. (Fodor, 1974, 1997; Mitchell, 2009; Kaiser, 2017; Mazzocchi, 2008; Batterman, 2001, 2005). So, it seems that there are plenty of systems that can satisfy the premise of DCA. Yet, one can still resist the conclusion of DCA by taking recourse to the optimistic counterargument.

The optimistic counterargument is an old one, going back to Hempel and Oppenheim (1948). The proponents of this counterargument are optimistic about the future of science, believing that the higher-level causal powers within special sciences are simply transient artifacts of the current incomplete state of science that will eventually be reduced to, and thus replaced by, lower-level explanations as science advances. From this optimistic perspective, Hempel and Oppenheim (1948) conclude that: “emergence of a characteristic is not an ontological trait inherent in some phenomena; rather it is indicative of the scope of our knowledge at a given time; thus it has no absolute, but a relative character; and what is emergent with respect to the theories available today may lose its emergent status tomorrow” (150–151).

The more extreme version of the argument contends that even our current state of science is at the verge of successfully reducing higher-level phenomena (Barwich, 2021; Bickle, 2006, 2020). For example, after discussion of some cases from neurobiology, Bickle (2006) writes:

[T]he result is a step toward a biophysical reduction of mind. Except for heuristic and pragmatic purposes, we will no longer need to speak of membrane potentials interacting with voltage-gated receptor proteins as a mechanism. The known biochemistry and biophysics ... will supersede the explanatory need to talk that way. The next step is to "intervene biophysically" with these newly discovered mechanisms and "track behaviorally." Successful examples will constitute mind-to-biophysics reductions, leaving molecular biology as a necessary heuristic but no longer the science for uncovering explanatory mechanisms. "Ruthless" reductionism grows positively merciless (432).

The counterargument need not be so “ruthless”. A weaker version of the argument in the form of an argument from ignorance will still be effective against metaphysical emergence. One could say that even if we accept that science has so far been unsuccessful at constructing a fully reductionistic theory of everything, and even if we are not sure that science will ever come up with such a theory, it is possible that it will. That possibility, so the thought goes, is enough to render DCA ineffective and prevent concluding metaphysical emergence from apparently irreducible phenomena and their higher-level distinctive causal powers.

All in all, any version of the optimistic argument is an important threat to various accounts of metaphysical emergence that are all, one way or another, inspired

and supported by claims of irreducibility of higher-level phenomena in special sciences. In fact, this counterargument has already forced the emergentists to retreat on a previous occasion. It was the wondrous achievements of science during the twentieth century and the alleged reduction of chemistry to quantum physics that resulted in the fall of the British emergentism of mid-19<sup>th</sup> and early twentieth centuries (McLaughlin, 1992). Those discoveries showed that the scientific phenomena that were commonly cited by the emergentists as irreducible examples were in fact reducible and, thus, proved the emergentists wrong, or at least so the anti-emergentists see the matter.

However, the following discussions aim to show the non-reductionistic face of modern science that is not compatible with the optimistic counterargument. I show that in many cases modern science does not pursue more and more reduction. On the contrary, it takes a holistic non-reductionistic approach. I argue this gives us evidence that the future of science would not necessarily be reductionistic and the irreducible emergent phenomena may not be transitory, but a permanent part of future science. After all, it seems that we should not be as optimistic about the possibility of a fully reductionistic future for science as the optimistic counterargument suggests, and we are not as ignorant about it as the weaker counterargument from ignorance implies.

Before embarking upon this line of reasoning, however, it is worth noting that there is also what I call the *pessimistic counterargument*. According to the pessimistic counterargument, it is impossible to come up with a completely reductionistic science, not because the world is populated by metaphysically emergent entities, but because of our inherent cognitive limitations, or certain computational constraints imposed on us by the structure of our world. The strongest version of the pessimistic counterargument can be found in writings of computational emergentists (Bedau, 2008; Huneman, 2008). According to computational emergentists, certain computational characters of the processes in our world, such as their so-called *computational irreducibility*, makes it theoretically impossible to come up with a fully reductionistic science. The non-reductionist approaches of science are merely a reflection of these computational constraints.

I have discussed these views in full detail and argued against computational accounts of emergence elsewhere (Tabatabaei Ghomi 2022). There I have tried to show that the conclusions of computational emergence do not follow from their underlying computational theories. Therefore, here I skip the discussion of those views and the associated pessimistic counterargument and focus on the optimistic counterargument.

## 5 From indecomposability to metaphysical emergence

In this section, I first explain how indecomposable systems show distinctive causal powers on the higher, systemic level and therefore, make good cases for DCA. I then argue, by analysing the heavy reliance of biological engineering on irrational methods, that biological systems are probably inherently indecomposable.



## 5.1 Indecomposability

Indecomposability means that the system does not lend itself to *decomposition*, a widely used strategy in special sciences such as biology (Bechtel & Richardson, 2010; Craver & Darden, 2013). So, to understand indecomposability, we need to first understand decomposition. In the process of decomposition, the overall function of a system, say a biological one, is decomposed into some smaller separate sub-functions called functional modules. For example, to explain protein biosynthesis, the whole general function is decomposed to modules such as transcription, translation, and post-translational modification. Each module is then localized to certain components of the biological system. In the case of protein biosynthesis these components are RNA polymerase, mRNAs, ribosomes, etc. These components, each performing a separate function, are supposed to interact with each other as puzzle pieces of an overall mechanism, and this mechanism produces the systemic functions such as synthesizing proteins.

Systems can be investigated by decomposition only on the assumption that they are inherently decomposable (Bechtel & Richardson, 2010; Rickles et al., 2007). Decomposable systems can be large and elaborate. Yet, their parts play specific identifiable functional roles, and the interactions between parts follow distinguishable rules. As a result, the function of a decomposable system can be reduced to the modular functions of its parts and their straightforward interactions. A car is an example of a complicated, yet decomposable system. Every car has about 30,000 parts that interact in elaborate ways. Yet, the manufacturer can tell you the exact function of each of these 30,000 parts and can describe how they work together to get the car going. The systemic function of the car is decomposable to its parts.

By contrast, systemic functions of indecomposable systems, commonly referred to as *complex* systems, are not decomposable to the parts and simple interactions. The dense and convoluted interactions and intertwined feedback and feed forward connections within these systems heavily influences the functions of their parts to the extent that the functions of the parts and their positions in the system become inseparable from one another. Consequently, one cannot describe standalone functions for each part. The parts get fused into an indecomposable system that can only be described as one whole unit rather than aggregation of separate modules. The systemic function can be ascribed only to the system as a whole, without being able to individuate the separate contribution of each part. As a result, the systemic causal powers of an indecomposable system are irreducible to anything simpler than the system itself. The system shows a causal profile that is irreducible and thus, distinguishable from the causal powers of its constituents. Such a system, therefore, satisfies the premise of DCA.

Over the past twenty years, many theorists have promoted the view that biological systems are indecomposable (Heng, 2017; Kaiser, 2017; Kauffman, 1993; Mazzocchi, 2008, 2011; Mikulecky, 2001; Plsek & Greenhalgh, 2001; Rickles et al., 2007; Shapiro, 2011; Walsh, 2015). Yet, the view of biological systems as truly indecomposable will not be established unless we address the optimistic counterargument in that context. For that purpose, let us switch from decomposition to recomposition, and go from biological discovery to biological engineering.

## 5.2 Biological engineering as recomposition

We can describe biological engineering as *recomposition* that follows *decomposition*. Decomposition is the reverse engineering of biological systems. The knowledge acquired by reverse-engineering sets the ground for *forward engineering*, or the recomposition of biological systems. Forward engineering of biological systems has a long history and has been tried at different levels, starting from biological parts, and going all the way up to engineering artificial life. The focus of this paper is on synthetic biology, the recent wave of biological engineering that rose around the millennium. Synthetic biology, at least in its idealized form, is the forward engineering of biological systems where the engineer deliberately assembles independent modules according to a pre-conceived plan to get a product with a desired function (Cameron et al., 2014; Lewens, 2013). Efforts to reverse-engineer biological systems gave rise to the view that cellular organisms are simply systems of discernible functional units similar to human-engineered machines (Cameron et al., 2014). Based on that view, scientists ventured to apply what they had learned from reverse engineering to forward engineer biological systems by assembling those functional units in new circuits. To those scientists' dismay, however, the attempts often failed and the designed systems did not behave as expected. Despite all the impressive recent advances, synthetic biological designs still fail to behave as expected, and the ideal engineering aspirations of the field remain far from realized (Cameron et al., 2014; Kwok, 2010).

One major problem facing biological engineering is the context-dependent behaviour of biological modules. When engineering non-biological systems, modules are usually well-characterized on their own and their functions do not change drastically irrespective of the system into which they are incorporated. A battery of a certain voltage, for example, provides more or less the same electrical power in all machines. The consistent behaviour of batteries allows us to simply take an AA battery from a drumming monkey and put it in our alarm clock. This is not the case, however, when it comes to biological modules. They behave differently from one system to another and it often takes considerable effort to exchange parts between biological systems (Lu et al., 2009). Biological parts behave differently even across systems as similar as various strains of a single species. For example, Bagh et al. built a very simple two-component system, a promoter gene that regulated the expression of a reporter protein (Bagh et al. 2008). This simple genetic circuit was put into four different strains of a single species, *E. coli*, and the expression of the reporter protein was followed. The level of protein expression varied significantly across the four strains of *E. coli*, and the authors could not explain how the small genetic differences of the hosts resulted in these significant variations (Bagh et al. 2008). It is as if you put the same battery in four slightly different drumming monkeys and get four completely different voltages.

Even much smaller biological units show significant sensitivity to much subtler changes in their contexts. An example is the concept of *epistasis* between mutations. In the context of proteins, epistasis happens when the effect of some particular mutation on the structure or the function of a protein depends on the sequence within which the mutation is introduced. Because of epistasis, not only may the

effect of single mutations differ from sequence to sequence, but the combined effects of two or more simultaneous mutations may deviate from the sum of their individual effects. Epistasis links the effect of multiple mutations to one another. For example, in a study by Weinreich et al. 14 different biological systems showed epistatic links ranging between three to seven mutations (Weinreich et al., 2013), and there is evidence that even more mutations may form extended epistatic groups (Halabi et al., 2009; Rivoire et al., 2016). In extreme cases of epistasis, a mutation that promotes a desired function may completely change its nature and impede that function if introduced concurrently with some other mutation (Starr & Thornton, 2016).

One explanation for context-sensitivity of biological modules and the consequent failures of biological engineering is that biological systems are *indecomposable*. In what follows, I aim to support this explanation by entertaining a number of alternatives and showing that indecomposability is indeed the best explanation.

### 5.3 Failures of rational biological engineering and the recourse to irrational methods

There can be three possible reasons for failures of synthetic biology. The first is an incomplete or wrong decomposition of the relevant biological systems that results in failure of following recomposition attempts. The second are practical limitations in realizing the engineering designs. The third is the indecomposability of biological systems. Each of these reasons would elicit a specific kind of reaction by biological engineers. By looking at the reaction of the engineers, I will infer the underlying reason for the failures.

Let us begin by the first possible reason for failures of biological engineering, which is the incomplete or wrong decomposition resulting in unsuccessful recomposition. This explanation is consistent with the optimistic counterargument and the argument from ignorance discussed above. Therefore, I analyse it in more details to show its infeasibility, at least as the sole, or the most important, explanation for failures of biological engineering. According to this explanation, biological engineers have missed some parts or drawn a wrong interaction map in the decomposition step and, consequently, their resulting recomposition is wrong or incomplete. It is the biological engineers who are to blame and not the method of decomposition. Decomposition is an appropriate method, so the thought goes, even though practitioners may fail to perform it properly.

If this is the case, failures of decomposition can indeed be fruitful as they result in what I call a *productive cycle*. Due to failure in recomposition, biologists go back and re-examine their decomposition of the system and come up with a revised decomposition that gives them a more accurate understanding of the system. They then test this new decomposition by another round of recomposition. In this way, recomposition provides a test platform to check if the proposed decomposition is accurate and complete. Biologists' understanding of the biological system improves through iterative cycles of decomposition-recomposition until they eventually get it right. As plausible the productive cycle model might look on paper, it does not fit what we observe happening in the practice of synthetic biology.

The first attempts at synthetic biology were two genetic circuits published in early 2000 by Collins' group, and Elowitz and Leibler, both of which concerned genetic circuits designed to induce certain desired functions into their host cells (Cameron et al., 2014). Collins' group designed a genetic circuit based on a natural genetic switch observed in bacteriophage  $\lambda$  that made its host cells toggle between two gene expression states (Gardner et al., 2000; Khalil & Collins, 2010). Elowitz and Leibler designed a circuit based on circadian oscillatory circuits observed in cyanobacteria that made the host show gene expression oscillation (Elowitz & Leibler, 2000; Khalil & Collins, 2010). The motivation behind these works was to reassemble natural modules and engineer an artificial biological system based on a pre-thought scheme. In both cases, however, researchers encountered considerable unexplainable noise, and contrary to their initial aspirations, had to rely not on pre-thought design, but on trial and error to get the final system. Consider the circuit developed by Collins' group. Roughly, the cells were expressing gene A, and a signal was supposed to turn off expression of gene A and prompt cells to express gene B. But the cells kept expressing gene A, and it took Collins group three years of tweaking to make this simple system work. After these three years no major parts were added to the design, nor the circuit was rewired. The understanding of the original natural system in bacteriophage  $\lambda$  also remained the same. The two gene promoters used had to be balanced against each other simply by trial and error (Kwok, 2010).

The unpredictability and inexplicable failure of biological designs haunted the field from the early days, led to a heavy reliance on trial and error in synthetic biology, and somewhat dulled the initial engineering enthusiasm (Cameron et al., 2014). Synthetic biology has advanced over recent years and better-characterized parts are found and more elaborate systems are built (Khalil & Collins, 2010; Lu et al., 2009). The problem of unpredictability of systemic behaviour, however, still poses a significant challenge to the field (Lu et al., 2009). Researchers have realized that even their well-characterized parts do not function as they think, and even their simple circuits do not behave as expected. The response to these failures was barely revisiting the decomposition of the systems to find missing parts or wrong arrangement maps and coming up with a new aforesought design. Rather, like the pioneering cases, subsequent synthetic biologists took recourse to trial and error. In technical terms, they reacted by shifting from the so-called *rational* methods to *irrational* methods.

Rational and irrational methods are two technical terms referring to two opposing research and development approaches and have nothing to do with philosophical rationality. What differentiates rational from irrational methods is whether the developer has a prospective understanding of how a system works on a mechanistic level (Lewens, 2013). If the developer possesses this understanding, she can rationally design a system with forethought, predict the behaviour of the resultant system, and fine-tune its performance accordingly.

But in fields such as biological engineering, rational methods often fail, and the developers turn into *irrational* methods. In irrational methods the researcher treats the system as a black box and relies on observations resulting from trial and error without necessarily having an explanation for them. In biological engineering, for example, she has to test many combinations of different biological modules hoping

to find the magic combination that shows the desired behaviour. She does not know how and why the system does what it does and therefore, once she finds one working system, she cannot touch its parts or modify its behaviour by rational re-design. To make any modifications in the system's behaviour she needs new rounds of trial and error.

One possible explanation for this turn towards irrational methods inspired by the optimistic view is that the developer does not yet understand the system on a mechanistic level and does not yet know how each part works and how different parts interact, and this is why she cannot design the system with forethought. This surely is the explanation behind many cases where rational methods fail and developers turn into irrational alternatives. However, if this is the only reason that rational methods fail and irrational methods are employed, we should observe a gradual shift from irrational methods towards rational ones as the relevant science and technology advance. I argue that in fields where we observe an opposite trend of more and more reliance on irrational methods, a passing gap in mechanistic knowledge does not tell the whole story behind the failures of rational approaches. I suggest that in such cases, inherent indecomposability of target systems is an important alternative explanation. The argument runs through the discussions of this section, and I present it in full and in formal format in Section 6.

The choice between rational and irrational methods is often not black-and-white. Biologists usually have partial knowledge of how their system works, and thus, adopt a partially rational, partially irrational approach. A synthetic biologist may have some idea about the type of parts, and the general design of the circuits that has the potential to generate the desired outcome. Using this partial knowledge, she limits her search space and starts with some tentative parts and initial sketches of the circuit. What converts this initial attempt to the final working system, however, are not multiple productive cycles, but are many rounds of trial and error. Even in those rare cases where biological engineers have been exceptionally successful in their initial designs, they needed irrational optimization to increase the performance of their systems up to an acceptable level. It is the case not only where biologists try to synthesize cellular circuits, but also when they try to develop smaller systems such as a single enzyme. Rationally designed enzymes, even the active ones, often do not show high enzymatic activity and are significantly inferior to their natural counterparts. Biologists have to use some irrational method such as artificial evolution to further optimize the rationally designed enzymes. Even a few cycles of artificial evolution might dramatically improve the performance of the designed enzymes (Golynskiy & Seelig, 2010). This improvement is usually about 100-fold increase in activity, and in some cases can be as dramatic as 10,000-fold increase or more (Khersonsky et al., 2010). Irrational methods are indispensable steps of synthetic biology development, and it is expected that they will remain so (Cameron et al., 2014).

The shift from rational to irrational approaches is manifest not just in the experimental side of biological engineering, but also in the computational side. Starting around 2000s, deep learning methods have become more and more widely used to analyse large and complex biological data (Tang et al., 2019) and parallelly, their application has also grown in various sorts of biological engineering. Protein science is a telling example where deep learning methods are

growingly and successfully implemented. What is eye-catching is the dramatic success of deep learning methods in tasks such as protein structure prediction that has long been a daunting challenge for the classic approaches (AlQuraishi, 2019, 2020). Another interesting observation is the success of these methods in prediction of systemic and holistic characters of proteins such as their solubility (J. Chen et al., 2021), or dynamics (Degiacomi, 2019). Also on the engineering side, we are observing a wave of recent studies that show the power of deep learning methods in protein engineering (Alley et al., 2019; Biswas et al., 2020; Shroff et al., 2020; Xu et al., 2020). Protein science is not an exception and deep learning is showing its promise in various fields of biology with important engineering applications (Ching et al., 2018; Jones et al., 2017). Just as one example, a deep learning method to predict gene expression levels outperformed conventional linear regression for 99.97% of the target genes tested (Y. Chen et al., 2016).

The technical term of *irrational method* is not usually applied to describe deep learning methods. Nonetheless, I think we can view the shift from traditional more interpretable methods of data analysis to much less interpretable deep learning methods as another way that biology is shifting towards irrational approaches. One of the most important caveats of deep learning methods is the so-called black-box problem (Mamoshina et al., 2016). Despite their predictive success, it is hard, sometimes impossible, to interpret these models and infer the underlying causal relations that result in the correlations captured by these models. Although there are some techniques to help make sense of deep learning models (Montavon et al., 2018), it is unlikely that one gets the kind of interpretability of more traditional machine learning methods, especially in the elaborate models used in biological cases. The black-box problem means that similar to irrational experimentation, in deep learning the engineers rely on the overall outcome without necessarily knowing the underlying mechanisms. They have a scientifically approved crystal ball that tells them the answers but provides little explanation.

In short, the method of development in biological engineering, in experimentation and data analysis alike, is very different from the productive cycle model. We see a constantly growing reliance on irrational methods with no sign that this trend is going to change in the future. Constant and growing recourse to irrational methods instead of the productive cycle model in response to synthetic biology failures shows that it is unlikely that the optimist response that ascribes failures of biological engineering to temporarily incomplete or wrong decompositions can sufficiently explain all those failures. Wrong decompositions can definitely share the blame, but they cannot be the whole story.

This brings us to the second practical explanation that ascribes failures of biological engineering to technical limitations. The practical explanation suggests that the failure in synthetic biology developments and the following recourse to irrational methods is due to technical limitations in realising the intended designs. The idea is that biologists know what parts should be used, and they know how those parts should ideally be assembled to engineer the intended system. Nonetheless, they cannot create that system because they cannot realize that assembly. They may not have the parts they need, or they may not be able to put the parts in

the necessary arrangement. They know what should be done, so the thought goes, but they cannot do it as their hands are tied by their technological limitations. To find a way around those limitations, they have to rely on trial and error.

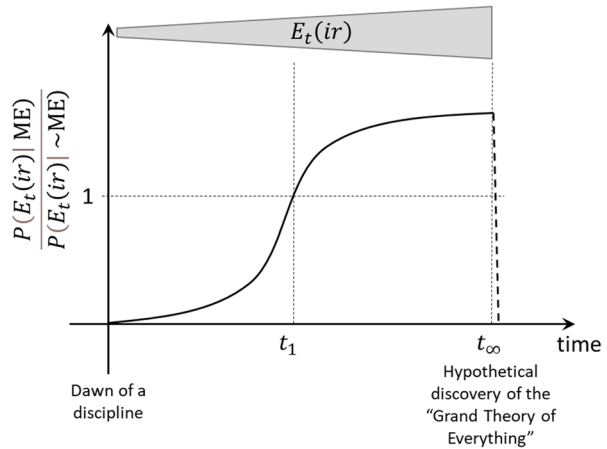
No doubt, this can be the reason behind some instances of failed synthetic biology development. But it does not capture the whole problem. There are many cases where synthetic biologists have all the parts they want, and they are able to put those parts in the arrangement they are aiming at and yet, their systems do not behave as expected. Actually, in many cases combinatorial methods are used to test not one, but hundreds, or even thousands of different combinations hoping to find the one combination that works (Khalil & Collins, 2010; Lewens, 2013). In such cases, biological engineers have little problem assembling a wide range of parts, in a wide range of ways. If they could find their systems by rational approaches, they would directly pick the working system without accepting the burden of testing many others. But they cannot, and they have to rely on trial and error. Therefore, the second, practical explanation also cannot be the whole story, and this takes us to the third, remaining explanation, which is the indecomposability of biological systems.

Indecomposability nicely explains the failures of biological engineering and the subsequent recourse to irrational methods. Because the functions of parts are under heavy influence of their encompassing indecomposable system, analyses of their functions in isolation or in another system tell very little about their function within the domain of the target system. This denies the biological engineer a priori knowledge of how the parts would work within the target system and consequently, prevents her from coming up with an a priori design. The engineer has to try different parts within the very context of the target system until she finds a working combination. As touching any of the parts may change the systemic state and subsequently affect how the other parts behave, different parts should be optimized simultaneously. These constraints leave the engineer with no choice but to use irrational methods of development that allow choosing the parts within the context of the target system and optimizing the system in its entirety.

Indecomposability also explains why deep learning methods perform so well in biological contexts. The independent variables produced by some of the biologically successful deep learning methods are generated by a non-linear combination of different apparently independent and unrelated variables. Such a combination of seemingly separate variables seems to be the appropriate mathematical description of an indecomposable system in which several apparently separate actors get combined into intertwined holistic units. The success of deep learning methods in biological contexts, therefore, hints at the indecomposability of the modelled systems.

Rational methods provide insight into the underlying mechanisms and map out a more straightforward path to developing the desired systems. Scientists, therefore, often prefer to stick with rational methods. Yet, when it comes to investigating and developing indecomposable systems, they have no choice but to resort to irrational methods. Wherever scientists opt for irrational over rational methods, we should suspect that they are forced into it because their subject of study is indecomposable. The prevalent, continuous, and growing application of irrational methods in biological engineering, therefore, provides evidence that biological systems are indecomposable systems. As such, they are unlikely to be decomposed in the future.

**Fig. 2** Confirmation of meta-physical emergence by evidence of irrationality over time



## 6 The optimistic counter-argument and the evidence coming from irrational methods

The reductionists who endorse the optimistic counterargument might recognise, or even promote using irrational approaches.<sup>3</sup> However, as supporters of the optimistic counterargument, these reductionists might attribute the heavy reliance on irrational methods to an immature understanding of the system under study or development, or technical limitations. If this is the case, then irrational methods are expected to give way to rational approaches as the relevant field of science and technology matures. The question, however, is how much weight one should give to this optimistic picture of the future. We might never be able to completely prove or reject the possibility of this optimistic future. But we could, and we should, adjust our estimates of its possibility based on available evidence, particularly the evidence coming from the current practice of science and its ongoing trajectory. Our views about the future of science should be based on the path that it has taken so far and where it seems to be heading now from its current point. In what follows, I propose how we should adjust our predictions about the future of science, and correspondingly, our metaphysical views in light of the evidence coming from the current irrational practices within science.

We saw that reliance on irrational methods provides some evidence in support of indecomposability of the systems under investigation, impossibility of the optimistic prediction, and henceforth, the existence of metaphysical emergence. But this is only one piece of defeasible evidence, and so we should not rush into conclusions. We should inspect the course of maturation of a discipline of science and evaluate the use of rational methods,  $r$ , compared to irrational methods,  $ir$ , as the discipline progresses. Suppose that we observe that irrational methods are gaining more and

<sup>3</sup> Bickle, for instance, whom I cited as a prime reductionist, highlights the importance of these methods (Bickle, 2019).



more prominence. Let us call this the evidence of irrationality, or  $E(ir)$ , and denote the amount of this evidence at time  $t$  by  $E_t(ir)$ . The grey cone on top of Fig. 2 denotes increasing  $E_t(ir)$ .

We can assess two alternative hypotheses in light of this evidence. Either the subjects of that discipline are decomposable systems that are yet to be decomposed correctly (Dec), or we are dealing with inherently indecomposable systems that can be investigated solely by irrational methods (InDec). The likelihood ratio for these two alternative hypotheses is:

$$\frac{P(E_t(ir)|\text{InDec})}{P(E_t(ir)|\text{Dec})}$$

Following DCA, InDec concludes in the existence of metaphysical emergence (ME), while Dec lends support to generative atomism and the non-existence of metaphysical emergence ( $\sim$  ME). Thus, the above likelihood ratio positively correlates with the following likelihood ratio:

$$\frac{P(E_t(ir)|\text{ME})}{P(E_t(ir)|\sim \text{ME})}$$

Figure 2 summarises the way the above likelihood ratio changes over time. At the dawn of a discipline, the discipline is still young and immature, and the likelihood ratio is less than one and in favour of  $\sim$  ME. However, as the discipline matures, if irrational methods become increasingly prominent, at some point of time  $t_1$ , the likelihood ratio will eventually tilt in favour of ME. Even then the evidence does not confer certainty and may be defeated by future evidence. Yet for the time being, ME would be warranted. I hope that the detailed empirical discussions above have shown that we have passed  $t_1$  for many biological systems.

From the Bayesian perspective, a high likelihood in favour of a hypotheses does not necessarily mean that the hypothesis has high probability. The prior probability might be too low to begin with. Thus, one who strongly adheres to the metaphysics of generative atomism might accept that heavy reliance on irrational methods provides good evidence in favour of ME, and yet reject ME by assigning it a very low prior probability. But why should one adhere so strongly to generative atomism? As Humphreys (2016b) correctly points out, generative atomism owes its popularity to some alleged scientific successes in reducing emergent phenomena. A scientifically minded philosopher who has accepted generative atomism based on evidence from science should be ready to give it up if further scientific evidence speaks against it.

What if at some distant future researchers finally find the Grand Theory of Everything that reduces science to fundamental physics? Arguably, our historical evidence cannot exclude the logical possibility of such future discovery. If this happens, then that Grand Theory would explain away indecomposibility and irreducible systemic causal powers, establish generative atomism, and wash away metaphysical emergence. But as Mitchell (2012) puts it, “[t]o assume in an argument what we might know at ‘the end of science,’ ... is to ignore the facts of the history of science and the state of current science.” Until we reach “the end of science,” we should

take Cartwright's advice and make sure that our metaphysics walks hand in hand with our methods (Cartwright, 2007). As long as scientists of a mature discipline are obliged to use irrational methods, we have good evidence in favour of the existence of metaphysical emergence within the phenomena investigated in that discipline.

Here, we are in one of those situations where absence of evidence can be evidence of absence. Sober (2009) suggests that in cases where it is theoretically possible to observe some evidence and we have looked hard to observe it, then absence of evidence can be evidence of absence. In mature disciplines where many generations of scientists have tried hard to develop a reductionistic Theory of Everything, lack of such a theory and a growing reliance on methods that take the discipline further away from such a theory is evidence that such a theory may not exist.

## 7 Conclusions

The optimistic counterargument, the view that in some future time science will reduce everything to fundamental physics, works against irreducible higher-level causes acting as convincing evidence for the existence of metaphysical emergence. Those causes seem to be irreducible, so the counterargument goes, but science will eventually reduce them to lower-level causes, or at least it is probable that this will happen. I analysed synthetic biology as an example and showed that the evidence from heavy reliance on irrational methods in that discipline speaks against this optimistic view in biology. I generalised that such optimistic predictions lose their warrant for any mature discipline which relies continually and expansively on irrational methods. I concluded that recourse to irrational methods is a probabilistic marker that points to indecomposability and, therefore, metaphysical emergence. In summary, I showed that *irrationality suggests indecomposability, and indecomposability implies emergence*.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alexander, S. (1920). *Space, time, and deity*. Dover.
- Alley, E. C., Khimulya, G., Biswas, S., AlQuraishi, M., & Church, G. M. (2019). Unified rational protein engineering with sequence-based deep representation learning. *Nature Methods*, *16*(12), 1315–1322.
- AlQuraishi, M. (2019). AlphaFold at CASP13. *Bioinformatics*, *35*(22), 4862–4865.
- AlQuraishi, M. (2020). Protein-structure prediction gets real. *Nature*, *577*(7792), 627–628.

- Bagh, S., Mazumder, M., Velauthapillai, T., Sardana, V., Dong, G. Q., Movva, A. B., Lim, L. H., & Mcmillen, D. R. (2008). Plasmid-borne prokaryotic gene expression : Sources of variability and quantitative system characterization. 1–12. <https://doi.org/10.1103/PhysRevE.77.021919>
- Barwich, A.-S. (2021). Imaging the living brain: An argument for ruthless reductionism from olfactory neurobiology. *Journal of Theoretical Biology*, 512, 110560.
- Batterman, R. W. (2001). *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence*. Oxford University Press.
- Batterman, R. W. (2005). Response to Belot's "whose devil? Which details? *Philosophy of Science*, 72, 154–163.
- Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. MIT Press.
- Bedau, M. A. (1997). Weak emergence. *Noûs*, 31(s11), 375–399.
- Bedau, M. A. (2008). Is weak emergence just in the mind? *Minds and Machines*, 18(4), 443–459.
- Bickle, J. (2006). Reducing mind to molecular pathways: Explicating the reductionism implicit in current cellular and molecular neuroscience. *Synthese*, 151(3), 411–434.
- Bickle, J. (2019). Linking mind to molecular pathways: The role of experiment tools. *Axiomathes*, 29(6), 577–597.
- Bickle, J. (2020). Laser lights and designer drugs: New techniques for descending levels of mechanisms 'in a single bound'? *Topics in Cognitive Science*, 12(4), 1241–1256.
- Biswas, S., Khimulya, G., Alley, E. C., Esvelt, K. M., & Church, G. M. (2020). Low-N protein engineering with data-efficient deep learning. *BioRxiv*.
- Cameron, D. E., Bashor, C. J., & Collins, J. J. (2014). A brief history of synthetic biology. *Nature Reviews Microbiology*, 12(5), 381–390. <https://doi.org/10.1038/nrmicro3239>
- Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge University Press.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Chalmers, D. J. (2008). Strong and weak emergence. In P. Clayton & P. Davies (Eds.), *The re-emergence of emergence: The emergentist hypothesis from science to religion* (pp. 245–54). Oxford University Press.
- Chen, Y., Li, Yi., Narayan, R., Subramanian, A., & Xie, X. (2016). Gene expression inference with deep learning. *Bioinformatics*, 32(12), 1832–1839.
- Chen, J., Zheng, S., Zhao, H., & Yang, Y. (2021). Structure-aware protein solubility prediction from sequence through graph convolutional network and predicted contact map. *Journal of Cheminformatics*, 13(1), 1–10.
- Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of the Royal Society Interface*, 15(141), 20170387.
- Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.
- Degiacomi, M. T. (2019). Coupling molecular dynamics and deep learning to mine protein conformational space. *Structure*, 27(6), 1034–1040.
- Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403, 335–338.
- Fodor, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese*, 28, 97–115.
- Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Philosophical Perspectives*, 11, 149–163.
- Forrest, P. (2020). The identity of indiscernibles. In EN Zalta (Ed.), *The [Stanford] Encyclopedia of Philosophy*. {W}inter 2. Metaphysics Research Lab, Stanford University.
- Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403, 339–342.
- Golynskiy, M. V., & Seelig, B. (2010). De novo enzymes : From computational design to mRNA display. *Trends in Biotechnology*, 28, 340–345. <https://doi.org/10.1016/j.tibtech.2010.04.003>
- Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press.
- Halabi, N., Rivoire, O., Leibler, S., & Ranganathan, R. (2009). Protein sectors : Evolutionary units of three-dimensional structure. *Cell*, 138(4), 774–786. <https://doi.org/10.1016/j.cell.2009.07.038>

- Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2), 135–175.
- Heng, H. H. Q. (2017). The conflict between complex systems and reductionism. *Journal of American Medical Association*, 300(13), 1580–1581.
- Horgan, T. (1989). Mental quausation. *Philosophical Perspectives*, 3, 47–76.
- Humphreys, P. (1997). How properties emerge. *Philosophy of Science*, 64(1), 1–17.
- Humphreys, P. (2016a). Basic features of emergence. In *Emergence: A philosophical account*. Oxford University Press.
- Humphreys, P. (2016b). Ontological emergence. In *Emergence: A philosophical account*. Oxford university press.
- Huneman, P. (2008). Emergence made ontological? Computational versus combinatorial approaches. *Philosophy of Science*, 75(5), 595–607.
- Jones, W., Alasoo, K., Fishman, D., & Parts, L. (2017). Computational biology: Deep learning. *Emerging Topics in Life Sciences*, 1(3), 257–274.
- Kaiser, M. I. (2017). The limits of reductionism in the life sciences. *History and Philosophy of Life Sciences*, 33(4), 453–476.
- Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press.
- Khalil, A. S., & Collins, J. J. (2010). Synthetic biology : Applications come of age. *Nature Publishing Group*, 11(5), 367–379. <https://doi.org/10.1038/nrg2775>
- Khersonsky, O., Röthlisberger, D., Dym, O., Albeck, S., Jackson, C. J., Baker, D., & Tawfik, D. S. (2010). Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the Ke07 series. *Journal of Molecular Biology*, 396(4), 1025–1042. <https://doi.org/10.1016/j.jmb.2009.12.031>
- Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT press.
- Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, 95(1), 3–36.
- Kwok, R. (2010). Five hard truths for synthetic biology: Can engineering approaches tame the complexity of living systems? *Nature*, 463(7279), 288–291.
- Lewens, T. (2013). From Bricolage to BioBricks™ : Synthetic biology and rational design. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 44, 641–648. <https://doi.org/10.1016/j.shpsc.2013.05.011>
- Lu, T. K., Khalil, A. S., & Collins, J. J. (2009). Next-generation synthetic gene networks. *Nature Biotechnology*, 27(12), 1139–1150. <https://doi.org/10.1038/nbt.1591>
- Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Molecular Pharmaceutics*, 13(5), 1445–1454.
- Mazzocchi, F. (2008). Complexity in biology, exceeding the limits of reductionism and determinism using complexity theory. *EMBO Reports*, 9(1), 10–14.
- Mazzocchi, F. (2011). The limits of reductionism in biology: What alternatives. *Electronic Journal of Philosophy*, 11, 1–19.
- McLaughlin, B. P. (1992). The rise and fall of british emergentism. In A. Beckeran, H. Flohr, & J. Kim (Eds.), *Emergence or reduction?: Essays on the prospects of nonreductive physicalism*. Walter de Gruyter.
- Mikulecky, D. C. (2001). The emergence of complexity : Science coming of age or science growing old ? *Computer and Chemistry*, 25, 341–348.
- Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.
- Mitchell, S. D. (2012). Emergence: Logical, functional and dynamical. *Synthese*, 185(2), 171–186.
- Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1–15.
- Nagel, T. (2012). *Mind and cosmos: Why the materialist Neo-Darwinian conception of nature is almost certainly false*. Oxford University Press.
- O'Connor, T., & Wong, H. Y. (2005). The metaphysics of emergence. *Noûs*, 39(4), 658–678.
- Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis.
- Orengo, C. A., & Thornton, J. M. (2005). Protein families and their evolution—a structural perspective. *Annual Review of Biochemistry*, 74, 867–900.
- Plsek, P. E., & Greenhalgh, T. (2001). The challenge of complexity in health care. *BMJ*, 323, 625–628.

- Rickles, D., Hawe, P., & Shiell, A. (2007). A simple guide to chaos and complexity. *Journal of Epidemiology and Community Health*, *61*, 933–937. <https://doi.org/10.1136/jech.2006.054254>
- Rivoire, O., Reynolds, K. A., & Ranganathan, R. (2016). Evolution-based functional decomposition of proteins, 1–26. <https://doi.org/10.1371/journal.pcbi.1004817>.
- Shapiro, J. A. (2011). *Evolution: A view from the 21st century perspective*. FT Press Science.
- Shroff, R., Cole, A. W., Diaz, D. J., Morrow, B. R., Donnell, I., Annapareddy, A., Gollihar, J., Ellington, A. D., & Thyer, R. (2020). Discovery of novel gain-of-function mutations guided by structure-based deep learning. *ACS Synthetic Biology*, *9*(11), 2927–2935.
- Sober, E. (2009). Absence of evidence and evidence of absence: Evidential transitivity in connection with fossils, fishing, fine-tuning, and firing squads. *Philosophical Studies*, *143*(1), 63–90.
- Starr, T. N., & Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Science*, *25*, 1204–1218. <https://doi.org/10.1002/pro.2897>
- Tabatabaei Ghomi, H. (2022). Setting the demons loose: Computational irreducibility does not guarantee unpredictability or emergence. *Philosophy of Science*, *89*(4), 761–783. <https://doi.org/10.1017/psa.2022.5>
- Tang, B., Pan, Z., Yin, K., & Khateeb, A. (2019). Recent advances of deep learning in bioinformatics and computational biology. *Frontiers in Genetics*, *10*, 214.
- Van Cleve, J. (1990). Mind–dust or magic? Panpsychism versus emergence. *Philosophical Perspectives*, *4*, 215–226.
- Walsh, D. M. (2015). *Organisms, agency, and evolution*. Cambridge University Press.
- Weinreich, D. M., Yinghong Lan, C., Wylie, S., & Heckendorn, R. B. (2013). Should evolutionary geneticists worry about higher-order epistasis? *Current Opinion in Genetics & Development*, *23*(6), 700–707. <https://doi.org/10.1016/j.gde.2013.10.007>
- Wilson, J. M. (2015). Metaphysical emergence: Weak and strong. In T. Bigaj & C. Wüthrich (Eds.), *Metaphysics in contemporary physics; Poznan studies in the philosophy of sciences and the humanities* (pp. 251–306). Brill.
- Wilson, J. M. (2021). *Metaphysical emergence*. Oxford University Press.
- Xu, Y., Verma, D., Sheridan, R. P., Liaw, A., Ma, J., Marshall, N. M., McIntosh, J., Sherer, E. C., Svetnik, V., & Johnston, J. M. (2020). Deep dive into machine learning models for protein engineering. *Journal of Chemical Information and Modeling*, *60*(6), 2773–2790.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.