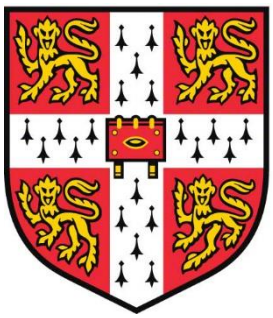
Chiffchaffs chirp and cherries are real: a scientifically informed defence of wholehearted emergentism



Hamed Tabatabaei Ghomi

Wolfson College
University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

December 2023

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared below and specified in the text. It is not substantially the same as any that I have submitted or is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted for any such degree, diploma, or other qualification at the University of Cambridge or any other University or similar institution.

- Chapter 3 is based on a collaboration with Dr Antonio Benítez Burraco

This dissertation does not exceed the word limit of 80,000 words set by the Degree Committee of the Department of History and Philosophy of Science, University of Cambridge.

Chiffchaffs chirp and cherries are real: a scientifically informed defence of wholehearted emergentism

Hamed Tabatabaei Ghomi

ABSTRACT

In this thesis, I provide a scientifically informed defence of metaphysical emergence, the idea that some systems are metaphysically distinct from their constituent parts and possess properties that are not reducible to their parts' properties. Moreover, I argue that metaphysical emergence is acceptable only if embraced wholeheartedly, along with all its ontological corollaries. Finally, I offer an ontology that suits metaphysical emergence.

I begin, in chapter 1, by rejecting a computational class of theories of emergence that I take to be the most important rivals to metaphysical emergence. These theories try to explain the irreducible properties of emergent phenomena by reference to their computational irreducibility. I show that computational irreducibility fails to fulfil its philosophical roles in these theories.

In chapter 2, I offer a practical argument in support of 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 an alternative reductionistic metaphysics.

In chapter 3, I analyse the emergentist research programme in linguistics as an example of the scientific application of the concept of emergence. The underlying theme of this chapter is that half-hearted emergentism is hopeless. I show that if one adopts some weak understandings of the concept of language emergence, the emergentist programme is not fundamentally different from the other non-emergentist research programmes in linguistics. On the other hand, if one adopts some stronger understandings of emergence then the programme would have a unique character, but at the cost of some philosophical corollaries that demand a fundamental revision of the emergentist programme in its present shape.

In chapter 4, I suggest that if one accepts metaphysical emergence, then one needs to revise one's ontological views as well. I compile the minimum set of ontological commitments necessary for maintaining the metaphysical and causal claims of structuralist accounts of metaphysical emergence. The set has three elements: (A) structural realism, (B) structural causation, and (C) the condition of downward percolation.

ACKNOWLEDGEMENTS

I am deeply indebted to my advisor, Jacob Stegenga, for his superb support, insightful guidance, and all-rounded training. Philosophy has been only one of the many things I have learnt from him, and I am so honoured to have him as my mentor.

I would like to thank my supervisor, Tim Lewens, for his instructive comments on this dissertation, and his general guidance and support over the years. Also, thanks to Antonio Benítez Burraco for a nice collaboration that enriched my dissertation, and to Mohammad Reza Mousavi, Hamed Ghasemieh, and Ali Haghi for their helpful comments on the first chapter. Thanks also to *Philosophy of Science*, *the European Journal for the Philosophy of Science*, and *Theoria* for publishing my first, second, and third chapters, and to their anonymous reviewers for their thoughtful comments. And thanks to my examiners, Denis Walsh and Neil Dewar, for their careful reading of my dissertation, fruitful discussions during the viva, and their written commentary. I feel proud remembering that my dissertation has passed their scrutiny.

I would like to extend my thanks to the staff and the students of the Department of History and Philosophy of Science at the University of Cambridge for making this department such an intellectually vibrant and socially comfortable centre of scholarship and education. Particularly, I want to acknowledge Adrian Erasmus, Zinhle Mncube, Victor Parchment, Elizabeth Seger, Cristian Larroulet Philippi, Adrià Segarra, Ina Jantgen, Sophia Crüwell, Benjamin Chin-Yee, Charlotte Zemmel, Oliver Holdsworth, Arthur Harris, Michaela Egli, Matthew Gummess, and Dominique Waissbluth for their friendship and camaraderie, and for many stimulating discussions and helpful comments. Thanks should also go to Robert Northcott, Richard Jennings, Misha Golynskiy, and Sergio Peisajovich for their heart-warming professional support over my PhD years. And I would also want to gratefully acknowledge all my fantastic friends in the Iranian community of Cambridge who enriched my years in this town with poetry, music, and spirituality. I cherish your friendship so dearly.

The culmination of my philosophical training at the University of Cambridge would have been impossible without the tremendous support and training I had received before coming to Cambridge. In this regard, I would like to extend my sincere gratitude to Cathy Gere, whose exceptional teaching, insightful guidance, and generous support played key roles in my acceptance to the University of Cambridge. I would also want to sincerely thank Matthias Steup, my MA thesis adviser, and the other professors who supported and guided me during my MA in philosophy at Purdue University. In particular, I want to acknowledge Daniel Smith for his kind support and

encouragements, and his enjoyable and inspiring classes on the philosophy of art. Thanks also to Markus Lill, who generously allowed me to complete my MA in philosophy while working on my first PhD in computational drug design under his supervision. And thanks to Ebrahim Azadegan, for his advice in one consultatory email exchange that was instrumental in setting the course of my philosophical academic journey.

I cannot even begin to express my thanks to my parents, Massi and Hossein Tabatabaei Ghomi. None of my accomplishments would have been possible without your nurturing, and I am eternally grateful for your unending support, love, and encouragement.

And my last and deepest gratitude goes to my love, Elaheh, who has been my safe refuge in difficulties, my fountain of hope in dejections, and my wise oracle in perplexities. I could not achieve this without your inexhaustible patience and endless sacrifice over many years. There are not enough words to express how much I appreciate you.

Spiritual thoughts have always been a beacon of hope and guidance for me. So, let me close my acknowledgements with a few words from Quran: This is of my Lord's bounty that He may try me, whether I am thankful or ungrateful. Whosoever gives thanks gives thanks only for his own soul's good …

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CA Cellular Automata

DCA The Distinctive Causal Powers Argument

MP Minimalist Programme

UG Universal Grammar

Introduction

The title of this dissertation consists of three parts which reflect the general motivation behind this dissertation and its overall structure. I will go through these three parts and introduce my motivations, moves, and conclusions. Let me begin with chiffchaffs and cherries. From a scientific perspective, there is a deep dichotomy between the world as we see it and the world as we know it. At least, this is what some teachers of science tell us. When walking in the woods, we see generous trees, joyful birds, hardworking insects, and soothing streams, and when looking inside ourselves, we 'see' bliss, calmness and fond memories associated with walking this same path. However, from what science has supposedly taught us, we know that the soothing stream is a mayhem of clashing H_2O molecules and that there are no generous trees, joyful birds, or hardworking insects – only chemical soups inside some physical pots. Moreover, bliss, calmness, and fond memories are merely some mindless neural wirings firing electrochemical signals caused by the balance of positive and negative ions suddenly changing on the two sides of the fatty layer. Indeed, all these phenomena, from memories to trees, just happen to be the way they are because some fundamental physical particles have the properties summarised in those weird-looking formulas of physics. In those formulas, so the thought goes, lies the ultimate metaphysical reality and causal power. In other words, the true metaphysical 'real' does not belong to the cherry but, rather, to its constituent atoms, and what truly causes the chirp is not the chiffchaff but the fundamental physical causal powers.

Supporters of metaphysical emergence, however, beg to differ. Metaphysical emergence is the idea that some systems, as a whole, are entities in their own right that are above and beyond their constituent parts and have their own powers and properties at the systemic level, which are distinct from those of the parts. Things like joyful birds, soothing streams and blissful thoughts are such metaphysically emergent systems, and, therefore, they are real entities with their own powers and properties, regardless of everything that we scientifically know about their constituent physical parts. Thus, according to metaphysical emergentists, chiffchaffs do cause chirps and cherry trees are indeed metaphysically real.

Let us move on to the second part: 'a scientifically informed defence of'. A scientifically informed defence of metaphysical emergence, such as the present piece, suggests that metaphysical emergence is not anti-scientific but scientifically supported. It suggests that there are metaphysically emergent systems, not *despite*, but *in light* of what we scientifically know about them.

The idea is that, at least in some disciplines, scientific knowledge and methodology as well as the history of science provide evidence that some systems are metaphysically emergent. On a closer look into the science of emergent phenomena, the world as we know it seems to be getting closer to the world as we see it.

The first two chapters of this dissertation are dedicated to such scientifically informed defence. They support the idea of metaphysical emergence in a scientifically informed way, though they do this in two opposite ways. Chapter 1 supports metaphysical emergence in a negative way by rejecting a rival and supposedly scientifically backed position, whereas Chapter 2 supports the idea positively by providing a supporting argument. Chapter 1 rejects theories of computational emergence that describe emergence as an epistemic limitation imposed on us by computational characteristics of emergent systems. I see these theories as one of the most important rivalling theories against metaphysical emergence. According to these theories, an emergent phenomenon results from a *computationally irreducible* process, and this computational characteristic makes it impossible to predict the outcome of these processes except via simulation. This computational character, so these theories suggest, makes emergent systems *look* irreducible to their parts. This apparent irreducibility is an epistemic limitation that is imposed on us by computational characteristics of the emergent systems and their governing mathematical theories, and, because of the mathematical nature of this epistemic limitation, we can never break through it. Nonetheless, this is only an epistemic limitation and, contrary to what metaphysical emergentists claim, it does not mean that the emergent systems have metaphysical reality, nor do they have higher level causal powers beyond those of their parts. However, via technical analyses, I show that computational irreducibility can establish the impossibility of prediction only with respect to maximum standards of precision. Moreover, by articulating the graded nature of prediction, I show that unpredictability to maximum standards is not equivalent to being unpredictable in general. I conclude that computational irreducibility fails to fulfil its assigned philosophical role in theories of computational emergence, and, therefore, these theories have no solid basis.

In Chapter 2, I offer a practical argument for metaphysical emergence. I show that, in some disciplines, science increasingly relies on so-called irrational methods as opposed to so-called rational ones. Rational methods are those that are based more on some a priori theory about the underlying mechanisms in the system under study, while irrational methods are those that rely more on a posteriori observations of the system without relying on a clear mechanistic model. I argue that the growing reliance on irrational scientific methods provides evidence that objects of science are indecomposable and, as such, can be better described by metaphysical emergence instead of prevalent reductionistic metaphysics. I show that a potential *optimistic* 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 providing detailed examples from biological engineering, but the conclusions are extendable beyond that discipline. In the end, I summarise the main message of the chapter in the following adage: irrationality suggests indecomposability, and indecomposability implies emergence.

Chapters 3 and 4 are dedicated to the third part: 'wholehearted emergentism'. My general aim in Chapter 3 is to show that half-hearted emergentism is hopeless, and, to make this general point, I conduct an in-depth discussion concerning the concept of emergence as used in linguistics. There is a research programme in linguistics that is founded on describing language as an emergent phenomenon. I clarify and analyse how the concept of emergence is deployed in this emergentist programme. Throughout this clarification and analysis, I repeatedly approach junctions where one could adopt a stronger or a weaker understanding of the concept of emergence. I show that on the one hand, if one adopts the weaker understanding of the concept of language emergence, the emergentist programme is not fundamentally different from the other non-emergentist research programmes in linguistics and thus can be abandoned. On the other hand, if one adopts the stronger understanding of emergence, then the programme would have a unique character but at the cost of some philosophical corollaries and methodological commitments that the emergentist linguists would seemingly want to avoid. I conclude that the emergentist programme, as it stands, should be abandoned – unless language emergentists are ready to wholeheartedly commit to emergence and, consequently, reshape the programme in both theory and methodology. I summarise the general message of the chapter in the following adage: half-hearted emergentism is expendable, and wholehearted emergentism is expensive.[1]

Chapter 4 primarily addresses wholehearted emergentists who choose to maintain their position despite its expensive price. In this chapter, I begin by clarifying a specific version of wholehearted structural metaphysical emergence in terms of two claims: a metaphysical claim and a causal claim. I formulate these claims based on "metaphysical indecomposability", aiming to spell out the way in which an emergent whole is distinct from its parts. Then, I take this version of metaphysical emergence for granted and ask how the ontology of the world should be so that its claims could hold. This argument direction is against the usual stream of argumentation in the literature concerning metaphysical emergence. The literature on metaphysical emergence is usually

[1] This chapter was developed in collaboration with Dr. Antonio Benítez Burraco (University of Seville). The general idea of analyzing the concept of emergence in linguistics came up in a conversation with Dr. Benítez Burraco following my talk at MetaScience Project seminar at University of Bristol in 2021. Dr. Benítez Burraco is a linguist, and has been instrumental in pointing me to the relevant literature and topics in linguistics. However, I have been responsible for the philosophical side of the project and have led writing the paper.

concerned with ontological, scientific, methodological and other arguments for or against metaphysical emergence. In these debates, the acceptance or rejection of metaphysical emergence is a conclusion, not a premise. I turn this common practice on its head by starting from metaphysical emergence as a premise and searching for the ontological picture of the world that fits this premise. In particular, I look for the minimum set of ontological commitments that are necessary for having a substantive and robust structural account of metaphysical emergence. I analyse a number of representative structuralist accounts of emergence, focusing on the relation of their background ontological assumptions and their strengths and weaknesses as theories of emergence. Eventually, I conclude with a list of necessary ontological commitments with three elements: structural realism, structural causation, and the condition of downward percolation.

Although the four chapters are different pieces of a cohesive general picture, they each address a specific question, face their own readership, and are written as standalone self-contained articles. In fact, Chapters 1, 2, and 3 have been published as articles in the *Philosophy of Science* (Tabatabaei Ghomi, 2022) the *European Journal for Philosophy of Science* (Tabatabaei Ghomi, 2023), and *Theoria* (Tabatabaei Ghomi & Benítez-Burraco, 2023). This style of writing was more or less compulsory given the current literature on emergence. The term 'emergence' (as well as the terminology around it such as 'weak', 'strong' and so forth) means many things depending on who uses the terms and in which context. In this dissertation, I discuss these concepts within a diverse set of contexts and debates. Chapter 1 discusses them in the context of computational theories of emergence, Chapter 2 in the context of biological sciences and scientific methodology, Chapter 3 in the context of linguistics, and Chapter 4 in the context of metaphysics. The diversity of the discussions and relevant literature has caused me to define (and redefine) terms and present (and re-present) arguments according to the specific language of the relevant literature and the specific needs of each debate. I trust that these hopefully non-redundant redefinitions and re-presentations would prevent any confusion that might ensue from the terminological and conceptual confusions in the literature on emergence.

My reader has an expedition ahead, not only through various philosophical discussions but also the detailed technical discussions of computer theory, biological engineering, and linguistics. As reflected in the title of my dissertation, I have tried to present a scientifically informed philosophical work. After all, when dealing with the conundrums of metaphysical emergence, we cannot afford to spare any intellectual resources. I hope that the debates and the conclusions to follow will be interesting and instructive for both scientifically minded philosophers, and philosophically minded scientists.

Chapter 1
Setting the demons loose: computational irreducibility does not guarantee unpredictability or emergence

ABSTRACT

A phenomenon resulting from a computationally irreducible (or computationally incompressible) process is supposedly unpredictable except via simulation. This notion of unpredictability has been deployed to formulate recent accounts of computational emergence. Via a technical analysis, I show that computational irreducibility can establish the impossibility of prediction only with respect to maximum standards of precision. By articulating the graded nature of prediction, I show that unpredictability to maximum standards is not equivalent to being unpredictable in general. I conclude that computational irreducibility fails to fulfil its assigned philosophical roles in theories of computational emergence.

1.1 INTRODUCTION

Predictability is a common area of interest and investigation for both scientists and philosophers. It is not surprising, therefore, that some scientific theories on predictability have made their way to philosophical discussions. In particular, some philosophers have used computer science theories to formulate accounts of emergence that inform critical debates such as the nature of mind, the autonomy of the special sciences, or the origin of biological novelty (Bedau, 1997; Humphreys, 2016c; Huneman, 2008, 2012). In this chapter, I use one of these computational theories, Wolfram's computational irreducibility (Wolfram, 2002), or as some call it, computational incompressibility (Bedau, 2008; Huneman, 2008), as a foil to show how these theories fail to fulfil some of their assigned philosophical roles.

There are two major types of emergence discussed in the literature, weak and strong, with two corresponding types of unpredictability. The weakly emergent are reducible to their fundamental bases and therefore are unpredictable only in practice, while the strongly emergent are irreducible and therefore are unpredictable in principle. There is a third type of emergence discussed in the recent literature that sits somewhere between weak and strong emergence. According to

computational emergence, or as I will call it, ontological weak emergence, emergent phenomena are reducible to their bases, yet because of their specific computational characters, namely their computational irreducibility, they are, as Bedau puts it, unpredictable in practice in principle.

Wolfram (2002) claims that from the computational perspective, the outcomes of computationally irreducible processes are unpredictable for any observer. Philosophers such as Bedau (1997) and Huneman (2008) have adopted this view to formulate ontological weak emergence, according to which emergent phenomena are unpredictable for any observer because of the computational irreducibility of the path leading to those phenomena. These computational accounts of emergence supposedly have the best of the two worlds of weak and strong emergence, and are doubly attractive for the scientifically minded. They retain the familiar scientific metaphysics of weak emergence while they underpin the unpredictability of the emergent with an ontological and scientifically backed explanation.

However, I show that computational irreducibility, no matter how it's interpreted, guarantees unpredictability only with maximal standards of accuracy and precision, while the scientifically and practically relevant concept of predictability is graded and submaximal. This creates a dilemma for supporters of computational emergence; they cannot have the ontological cake and eat it, too. If they cite computational irreducibility as their ontological guarantee of unpredictability of the emergent, they have to stick to the maximal standards of prediction. The unhappy consequence, however, would be that all natural phenomena would trivially turn out to be unpredictable and "emergent", and computational emergence would be indiscriminatory, trivial, and irrelevant to the intuitive concept of emergence and the practical puzzles it creates.

On the other hand, if they ease the standards of prediction, computational irreducibility can no longer guarantee unpredictability and computational formulations of emergence will be sullied by subjectivity and arbitrariness. Consequently, computational ontological weak emergence would not be much different and superior compared to simpler and more flexible non-ontological accounts of weak emergence. If successful, the arguments will take ontological weak emergence off the table, and we go back to the choice between weak and strong emergence.

1.2 COMPUTATIONAL IRREDUCIBILITY

Wolfram describes the causal or the computational path leading to a phenomenon as an algorithm that computes that phenomenon. The original algorithm generating a phenomenon is computationally irreducible if it is the shortest possible path to derive that phenomenon (Wolfram, 2002). As there is no alternative pathway to compute the resulting phenomenon faster than the original algorithm, the resulting phenomenon is unpredictable (Figure 1.1).

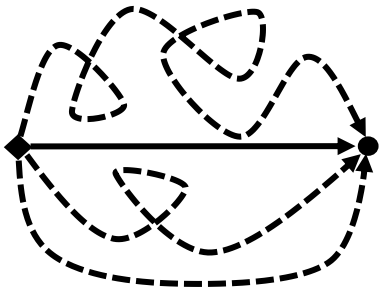


Figure 1.1 A computationally irreducible path to a phenomenon (solid line) is the shortest path to that phenomenon.

Wolfram presents the concept of computational irreducibility in the context of cellular automata, although the concept can be extended beyond that context. Cellular automata (CA) are lattices of cells called the *cell-space*, where each cell stores a value such as a colour. Starting from some initial values, the values of cells are updated in consecutive timesteps by some *updating rules*. At each timestep, the updating rule looks at a cell's own value and its neighbouring cells' values and determines the cell's value in the next timestep. Starting from some initial values and following the updating rules, some general patterns of values form on the cell-space as time goes by. The value of any single cell in each step is the *micro-dynamics* of a CA, and the overall patterns formed on the cell-space are the *macro-states*. Various combinations of cell shapes, dimensions of cell-space, cell values, and updating rules result in different CA (Berto & Tagliabue, 2017; Charbonneau, 2017).

Wolfram works with a particular setup of one-dimensional CA. Consider a chessboard with all white cells. Start by colouring some of the first top row cells black. In each iteration, move one row down and colour cells based on the colour of their neighbouring cells on the immediately above row following a set of updating rules. Each cell has three neighbours in the top row, and one, two or three of these neighbours might be black, allowing eight different colour combinations (Figure 1.2). An updating rule states whether a cell will be coloured black or not in each of these eight possible combinations. This means that we can have $2^8 = 256$ different updating rules. These

rules are numbered 0 to 255. Starting from some black and white cell configuration on the first row, different rules result in evolution of different general patterns on CA.

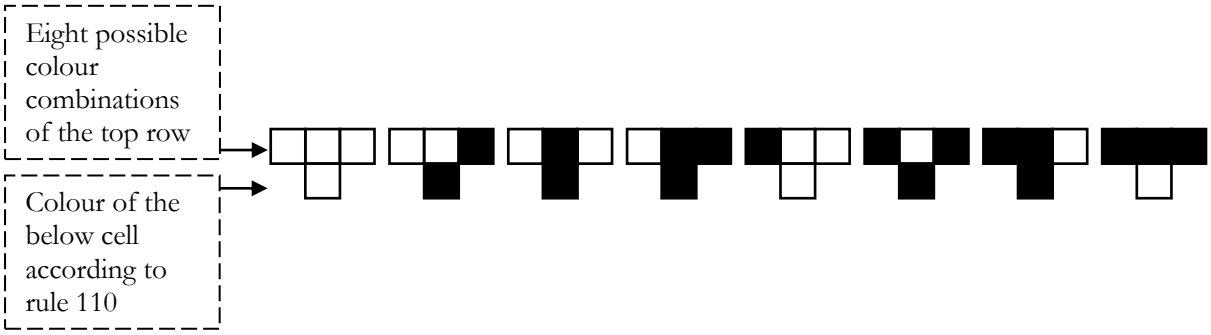


Figure 1.2 Overview of Wolfram's CA set up. Rule 110 is given as an example.

Wolfram categorizes these 256 rules into four classes (Figure 1.3). Class one comprises of those rules under which CA converges into a uniform final state, for example, all cells become black (e.g. rule 250). Class two are those rules that result in some simple periodic behaviour (e.g. rule 108), and class three are those that lead to random patterns (e.g. rule 90). Class four generates patterns that are a combination of randomness and periodicity (e.g. rule 110).

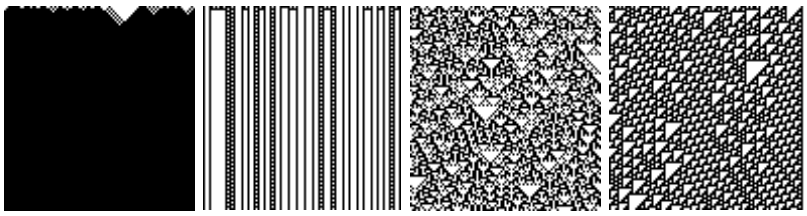


Figure 1.3 Examples of CA in each of Wolfram's four classes

Out of these four classes, the outcome of classes one and two are predictable using simple constant or oscillating predicting models. Using such predicting models, there is no need to simulate a class one or two CA in order to know its outcome. But when it comes to classes three and four, Wolfram claims that there are no predicting models and one has no alternative but to simulate the CA and watch what happens. Unlike the cases of classes one and two, there are no shortcuts, and the fastest way to know the outcome of the CA is going through the CA itself. These CA, therefore, are computationally irreducible (Wolfram, 2002).

Computational irreducibility is not limited to CA. Any algorithmic process would be computationally irreducible if the computationally most efficient way to know its outcome is running the algorithmic process itself. This is a mathematical hard lower bound and no observer whatsoever, not even a Laplacian demon, can pass it. The output of such an algorithmic process will be unpredictable because there is no way to compute its outcome faster than running the algorithm itself. In Wolfram's words (2002:739): "Whenever computational irreducibility exists in a system it means that in effect there can be no way to predict how the system will behave except by going through as many steps of computation as the evolution of the system itself".

The above apparently clear statement is in fact utterly ambiguous. Wolfram does not precisely clarify what he means by "as many steps of computation" and does not give a formal proof for his claim. Therefore, it remains unclear exactly in *what* sense and *why* a computationally irreducible process is *irreducible* after all. There are a few options to explain these *what* and *why*. I explore these options below and examine the type of unpredictability each can guarantee. The exploration will show that computational irreducibility can guarantee unpredictability only for *all cases*, in *infinite time*, or with *infinite precision*.

1.2.1 Computational irreducibility as a case of the halting problem

Wolfram deduces computational irreducibility from his *principle of computational equivalence*. This principle has multiple components, but the most important for the present discussion is the conjecture that any process, in nature or a computer, that does not converge to a constant or to an oscillating pattern is a *universal Turing machine* (M. Mitchell, 2009).

Turing machines are models of computers initially introduced by Alan Turing (1937) and now widely used in computer science. A simple Turing machine is composed of an infinite tape of individual cells as its memory and a head that points to one of these cells at a time. The head can read the cell it is pointing to, write on it, and move to the next neighbouring cell on either left or right. The machine can be in one of a set of predefined states. The behaviour of the head is dictated by a transition function based on the current state of the machine and the symbol read on the cell. The transition function determines what the head should write on the cell, the direction the head should move to and the state to which the machine should transit. A Turing machine begins at a "start" state and halts when it enters one of the "final" states. When the machine enters one of the final states, the computation finishes. But it is also possible that the machine never enters a final state and infinitely iterates over a loop of non-final states.

According to the Church-Turing thesis, any computable function is computable by some Turing machine. There are "universal" Turing machines that can simulate any other Turing machine and therefore, can compute any computable function. Cook has shown that Rule 110 CA is a universal Turing machine (Aaronson, 2002). This means that given the right initial conditions, a rule 110 CA can compute any computable function. Figuring out the right initial conditions for computing different functions with this CA is extremely hard. But it is in principle possible.

Although only a very limited number of CA are demonstrated to be universal Turing machines (Aaronson, 2002; Berlekamp et al., 1982; Rendell, 2011), Wolfram conjectures that any CA that does not converge to a constant or an oscillating pattern is a universal Turing machine. And because he sees natural processes as CA, he extends this claim to natural processes as well. Any natural process that does not result in a constant or oscillating behaviour, therefore, is a universal Turing machine, or so Wolfram claims. Both the initial conjecture and the extension to natural processes are open to objection. Nonetheless, let us grant both the conjecture and the extension and see where that would lead us.

Being a universal Turing machine subjects CA to the *halting problem*. The halting problem states that there is no single algorithm that can predict for any arbitrary input whether a universal Turing machine will halt. This can be one interpretation of computational irreducibility. Based on this interpretation, some CA are computationally irreducible and hence unpredictable because there is no general algorithm to predict if they will halt given some input. The only way to know whether they halt is to run the CA themselves (Ilachinski, 2001). For example, there is no single algorithm that can predict for any arbitrary initial pattern if Rule 110 will halt or continue to iterate indefinitely.

This unpredictability is not because of some technological limitation or lack of mathematical ingenuity. The halting problem dictates that halting or not halting of the computationally irreducible CA is unpredictable *in principle*. No technological breakthrough nor mathematical genius can ever devise a path around the halting problem (Yanofsky, 2016). Even a Laplacian demon would be stalled.

The conclusion drawn from the halting problem, however, is very restricted. The halting problem shows that no single algorithm can predict halting or not halting for all computations and for any arbitrary input. But the halting problem does not show that we cannot have multiple predicting algorithms that each correctly predicts halting or not halting for a broad subset of computations and inputs. Being subject to the halting problem is completely consistent with being

predictable in many instances. The halting problem only shows that we cannot predict the halting or not halting for all and every case.

In other words, the halting problem shows that there cannot be one single "Grand Algorithm of Everything" that predicts the fate of every natural process for all inputs. In this way, it seems that there might be a path from the halting problem to the impossibility of a "Theory of Everything". But the halting problem cannot show that particular natural processes are necessarily unpredictable by any theory. Scientists can appreciate the halting problem and yet go back to their benches and blackboards to make predictions.

I believe the recourse to the halting problem is the closest interpretation of computational irreducibility to what Wolfram has presented in his works. We saw that the unpredictability we can conclude from this interpretation is very restricted. But there are other ways to interpret computational irreducibility and those might be more successful in securing unpredictability. Let us explore those alternatives.

1.2.2 Algorithmic computational irreducibility in terms of algorithmic time complexity

Another way to understand Wolfram's irreducibility is to define it in terms of *algorithmic time complexity* (Huneman, 2008; Rucker, 2003; Zwirn, 2013; Zwirn & Delahaye, 2013). I explore a number of different versions of this approach in this section. But at their core and on a very rough sketch, all of those define a computationally irreducible algorithm as the one with the lowest (=best) algorithmic time complexity. It turns out that these formulations can guarantee unpredictability only over long times, and with infinite accuracy and precision.

Algorithmic time complexity is the computer science currency of speed and efficiency. In simple terms, it shows how the time needed for an algorithm to solve a problem scales with some measure of the input size. The time is estimated by the number of the elementary operations that an algorithm goes through to give the solution, and that number varies as a function of the input size. Time complexity is the *order* of this function which is expressed by notations such as the big O notation and is used to rank algorithms according to their efficiency.[2] The order shows how the number of operations scales with the input size. The lower the order, the more efficient the algorithm. Although time complexity is initially defined for algorithms, computer science usually associates time complexity with problems and not algorithms. The time complexity associated with

[2] Here I use the time complexity notations very loosely based on the general idea behind them and not their precise formalism. I stick with the big O notation as it is a weaker requirement compared to the small o notation.

a problem is the time complexity of the most efficient algorithm to solve it. Here, however, we need to zoom in, and discuss the time complexity of individual algorithms.

Here I devise a simple and familiar programming language to write the pseudo-codes. INPUT (X) means receiving an input and assigning its value to X. FOR (X) means repeating what is contained in the FOR-block X times. The FOR-block is determined by the brackets, the first of which is placed on the line immediately following the FOR command. OUTPUT (X) means finishing the programme and returning X as the output. Let us look at an example. Suppose you want an algorithm to compute the number of bacteria after *n* generations starting from a single bacterium. For simplicity, assume that no bacterium dies. In every generation, each bacterium divides and makes two daughter bacteria. This means that in each generation, one bacterium is added to the population per every bacterium in the previous generation. Here is the first algorithm (algorithm A) inspired by this observation:

Algorithm A:

```
INPUT (n)
current number of bacteria = 1
FOR (n)
        {
        number of bacteria in the previous generation = current number of bacteria
        FOR (number of bacteria in the previous generation)
                {
                current number of bacteria = current number of bacteria + 1
                }
        }
OUTPUT (current number of bacteria)
```

To determine the time complexity of this algorithm, we need to count the number of operations it performs for input *n*. The first FOR loop repeats the operations within its brackets for *n* generation. In each generation, the algorithm performs one addition per present bacteria. In the first generation, it does one addition, in the second, it does 2, in the third, it does 4, and so forth. The total number of additions performed by algorithm A, therefore, can be computed by the following formula:

$$1 + 2 + 4 + 8 + \cdots + 2^{n-1} = \sum_{i=0}^{n-1} 2^i = 2^n - 1$$

The function $2^n - 1$ is of order 2^n. Therefore, algorithm A has time complexity $O(2^n)$. It means that as the number of generations *n* increases, the number of operations needed to get the result, and consequently, the time spent to run the algorithm scales by the order of 2^n. This is an inefficient time complexity and there are algorithms of lower time complexities to solve this problem. For example, noting that the number of bacteria doubles in each generation, one can write algorithm B:

Algorithm B:
INPUT (n)
current number of bacteria = 1
FOR (n)
{
current number of bacteria = 2 × current number of bacteria
}
OUTPUT (current number of bacteria)

Algorithm B has time complexity $O(n)$, and this is significantly more efficient compared to algorithm A. If we understand computational reduction in terms of time complexity, then algorithm A is computationally *reducible* to algorithm B. We can use algorithm B to compute the result of algorithm A with a lower time complexity. And as algorithm B is more efficient than algorithm A, so the thought goes, we can predict the results of algorithm A using algorithm B.

There is an even more efficient algorithm to solve this problem. We know that starting from one bacterium, the number of bacteria after n generations equals 2^n. This translates to the following algorithm C:

Algorithm C:
INPUT (n)
current number of bacteria = 2^n
OUTPUT (current number of bacteria)

Note that 2^n is not an elementary operation and itself takes a few steps to compute. We have algorithms that compute 2^n with time complexity $O(\log(n))$.[3] Thus, algorithm C has time complexity $O(\log(n))$ that is a lower order compared to algorithm B's $O(n)$. Therefore, even algorithm B is reducible.

There is an important difference between algorithm C and the two previous algorithms A and B and that difference has inspired one interpretation of computational irreducibility. Contrary to algorithms A and B, there is no looping over generations in algorithm C. Algorithms A and B walk through the generations of cell division one by one, and they compute the number of bacteria for generations 1 to *n*-1 before they compute the number of bacteria in generation *n*. Algorithm C, on the other hand, does not go through this generation-by-generation path and jumps directly to generation *n* without computing the number of bacteria in the intervening generations. This is why algorithm C can be packaged as *a closed-form formula*, like a formula from physics:

$$\text{current number of bacteria} = 2^n$$

Rucker (2003) seems to interpret Wolfram's computational irreducibility as the impossibility of a closed-form formula. On this interpretation, a computationally irreducible algorithm is one the result of which cannot be computed by a closed-form formula and therefore, one needs to necessarily walk through the steps of the algorithm to know its results. Imagine that algorithm C or any other closed-form formula were not possible to solve our bacteria problem, and the only way to compute the result was to do the calculations generation-by-generation. Then, on Rucker's interpretation, the bacteria problem would have been computationally irreducible.

The example of the three algorithms above, however, shows that the mere impossibility of a closed-form formula is not enough to guarantee irreducibility and unpredictability. Even if algorithm C were not possible, algorithms A and B could be ranked according to their time complexity with the idea that the result of the more efficient algorithm B can predict the results of the less efficient algorithm A.

Zwirn and Delahaye (Zwirn & Delahaye, 2013) and Zwirn (Zwirn, 2013) formalize a more detailed definition for algorithmic computational irreducibility in terms of time complexity. Here I try to avoid the mathematics of Zwirn's account as much as possible and emphasize the general idea. The mathematical reader is referred to Zwirn's original paper for the exact formalism. Assume an algorithmic process *F(n)*, that computes f_n by going through f_1 to f_{n-1}. Starting from

[3] e.g. method of exponentiating by squaring

some initial input value for f_1, in each time step i, f_i receives the output of f_{i-1} and computes an output, which in turn will be the input for f_{i+1}.[4] On this setup, Zwirn requires two conditions for computational irreducibility of $F(n)$. The first condition requires that given n as input, a computationally irreducible F should compute f_n with the best possible time complexity. There should be no alternative algorithmic process G that computes f_n with lower time complexity. The second condition requires that any other alternative algorithmic process G that computes the same result f_n by an alternative function g should necessarily do so by going through steps g_1 to g_{n-1} where each g_i is an approximation of its corresponding step f_i. By "approximation" Zwrin means that one can compute f_i from g_i by a very short computation.[5]

Assume, just for the sake of argument, that the above algorithms A and B are the only possible algorithms to solve the bacteria problem. This (wrong) assumption guarantees that B has the lowest possible time complexity. It also ensures that the single alternative algorithm (algorithm A) computes the number of bacteria in generation n by going through steps 1 to n-1, and its value in each step can be easily transformed to the value computed by algorithm B (by B(n) = A(n)). B, therefore, would be computationally irreducible by Zwirn's definition.

Zwirn formulation has limited power to show unpredictability of the computationally irreducible. The reason is that having the same/lowest time complexity does not necessarily mean having the same/lowest absolute number of computations and the same/lowest run time in all conditions. For example, consider algorithm B' that solves the bacteria problem mentioned above as follows:

Algorithm B':

INPUT (n)

current number of bacteria = 1

FOR (n)

{

current number of bacteria = $\frac{100\times101}{\sum_{i=1}^{100} i}$×current number of bacteria

}

OUTPUT (current number of bacteria)

[4] Zwirn takes all f_i to be the same function f which runs iteratively, but I think this assumption is not necessary.

[5] Zwirn distinguishes *strong* computational irreducibility from computational irreducibility. The former satisfies the two conditions for any n, and the latter for infinitely many n. In this chapter I use computational irreducibility in the strong sense.

B' has the same time complexity as algorithm B, but it has a higher absolute number of computations and therefore, it runs slower than B (Figure 1.4). The difference between B and B' will eventually become insignificant as *n* increases to infinity. But in finite time, B can compute the results faster than B' and, therefore, can predict its outcome. Even more, B' has a better time complexity compared to algorithm A, but until *n*=10, A has a lower absolute number of computations and therefore, runs faster than B' and can predict its outcome (Figure 1.4).

It is only when *n* increases toward infinity that the difference between algorithms of the same time complexity diminishes to insignificance and the algorithms with the best time complexity are guaranteed to be faster than the alternative algorithms with higher time complexities. Therefore, computational irreducibility as defined by Zwirn guarantees unpredictability only toward the infinite time. For shorter times, a computationally irreducible algorithm might be predictable. Unpredictability based on Zwirn's definition is also limited by the possibility of heuristic solutions that I discuss at the end of this section.

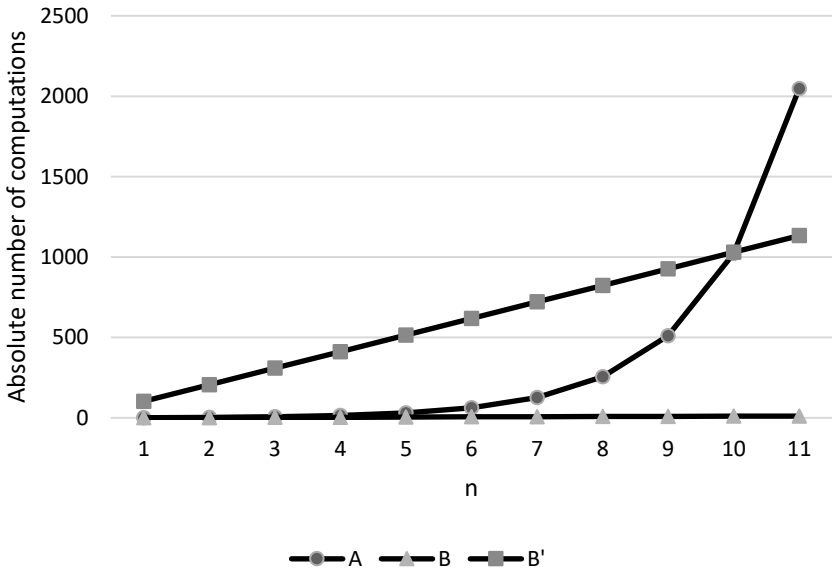


Figure 1.4 Comparison of the absolute number of computations in three algorithms. B and B' have the same time complexity. A has a higher time complexity compared to B and B'

On another interpretation, a process is computationally irreducible if predicting its outcome is an NP-hard problem (Buss et al., 1990; Huneman, 2008). Roughly speaking, NP hard problems are those for which we do not yet have a solution with polynomial time complexity ($O(n^{\alpha}), \alpha >$

1).[6] For any relatively large input size, it takes a prohibitively long time to solve a problem with time complexity of higher than polynomial, and therefore, NP-hard problems are practically unsolvable, or, as computer scientists call them, *intractable*. Intractability applies not just to us, the humans of the 21st century, but to any natural intelligence. For example, consider the Traveling Salesman, a classic NP-hard problem. The problem is finding the shortest path to visit each of n towns exactly once and come back to the starting town. A demon who can use all the atoms of the universe as processing units needs 10^{62} centuries to solve the Traveling Salesman problem for an input size of $n = 100$ (Yanofsky, 2016). As an NP-hard problem becomes intractable relatively quickly, we do not need to go toward infinite time for the prediction to become practically impossible.

Nevertheless, being NP-hard is also restricted in what it guarantees to be unpredictable. It is only the exact solution for an NP-hard problem that is intractable, and an approximate solution might be computable in much less time. In fact, NP-hard problems constitute a good portion of practical problems we are dealing with from everyday life to advanced science. Assigning wedding guests to seats in a way that friends share a table but foes do not (Lewis & Carroll, 2016), or constructing a phylogenetic tree (Habib & Stacho, 2013) are some examples. Computer scientists may not be able to provide exact solutions to these problems, but they have devised plenty of approximate solutions that are both fast and acceptably accurate. It does not even take an intelligent computer scientist to find fast approximate solutions for NP-hard problems. Zhu et al (2018) have shown that even as humble intelligence as an amoeba can find linear time approximate solutions for NP-hard problems.

Approximate solutions for NP-hard problems are examples of heuristic approaches that find approximate solutions for problems that are hard, or even impossible to solve. One particular approach that is specifically relevant to our discussion is coarse-graining of CA. Israeli and Goldenfeld (2006) have shown that computationally irreducible CA can be turned into reducible ones by coarse-graining. Coarse-graining of CA means combining some neighbouring cells into one cell. The process results in a lower-resolution and an imprecise re-description of a CA pattern. Israeli and Goldenfeld show that the patterns generated by a computationally irreducible class 4 rule (e.g. Rule 110) may be re-described as patterns of another rule of a lower class after coarse-graining (e.g. Rule 0) that are not computationally irreducible. This means that computational

[6] Buss et al (1990) use space complexity rather than time complexity. The two types of complexity are closely connected and as far as the arguments of this paper are concerned, there is no difference between the two.

irreducibility and its consequent unpredictability hold only for full precision CA. No matter how we define computational irreducibility, we lose it when we go to a coarse-grained view of CA.

In summary, different accounts of computational irreducibility based on algorithmic time complexity can guarantee unpredictability only for infinite time and infinite precision. Wolfram is aware of this problem, and he emphasizes this point in his previous papers. He writes (Wolfram 1984:424): "The large time limit of the entropy for class 3 and 4 cellular automata would then, in general, be non-computable: bounds on it could be given, but there could be no finite procedure to compute it to arbitrary precision."

But do we need arbitrary precision for predictability? And how far into the future should this "large time limit" be? In the next section I argue that to have a predictable phenomenon we do not need arbitrary precision and we do not need to predict until indefinitely long time limits.

1.3 PREDICTION IS GRADED

We generally understand prediction as describing a phenomenon before observing it, and we talk about "successful" or "failed" predictions, and accordingly, we say something is "predictable" or "unpredictable". These expressions of success, failure, predictability, and unpredictability represent prediction as a black and white concept that is either successful, or unsuccessful, possible, or not possible. This language usage hides the fact that prediction extends over a spectrum, and it is almost always not black or white, but grey.

This graded nature of prediction is well reflected in the science of prediction. There are fields like weather or political forecast where the success of prediction has been subject of heavy investigation. The metrics that these fields use to gauge the success of prediction rank various predictions on a range between complete success and complete failure. An example is the Brier score, one of the most widely used of these metrics (Brier, 1950). Brier originally formulated the score to measure the goodness of weather predictions, but later the score made its way to studies of socio-political predictions as well (Tetlock & Gardner, 2016). Brier score has the following formula:

$$B = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r}(p_{ij} - E_{ij})^2$$

The index j corresponds to each possible outcome, and r shows the total number of possible outcomes. For example, it might rain ($j = 1$) or not rain ($j = 2$) and in this case, $r = 2$. The index

i corresponds to an occasion of prediction, and n shows the total number of the predictions made. For example, we may predict whether it rains or not for ten days and in that case $i = 1, 2, \dots, 9, 10$ and $n = 10$. p_{ij} is the probability that the forecaster assigns to outcome j on occurrence i. E_{ij} indicates whether outcome j has happened on an occurrence i ($E_{ij} = 1$), or not ($E_{ij} = 0$). The best Brier score is zero, and the lower the Brier score, the better.

Suppose we have four weather forecasters. The Archangel who predicts all the rainy and not rainy days correctly and is always 100% confident about the predicted outcome. The Demon who mispredicts all days but is also 100% confident. The Bold forecaster who correctly predicts 60% of the rainy days, but she is 90% confident about her predictions. And the Cautious forecaster who also predicts 60% of rainy days correctly, but she is only 60% confident about her predictions. Table 1.1 shows the predictions of these four forecasters over ten consecutive rainy days and the corresponding Brier score.

Table 1.1 Some predictions and their corresponding Brier scores

	Day 1	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**	**Day 7**	**Day 8**	**Day 9**	**Day 10**	**Brier score**
Archangel	100% Rainy	100% Rainy	100% Rainy	100% Rainy	100% Rainy	100% Rainy	100% Rainy	100% Rainy	100% Rainy	100% Rainy	0
Demon	100% Sunny	100% Sunny	100% Sunny	100% Sunny	100% Sunny	100% Sunny	100% Sunny	100% Sunny	100% Sunny	100% Sunny	2
Cautious	60% Sunny	60% Rainy	60% Rainy	60% Rainy	60% Sunny	60% Rainy	60% Sunny	60% Rainy	60% Sunny	60% Rainy	0.48
Bold	90% Sunny	90% Rainy	90% Rainy	90% Rainy	90% Sunny	90% Rainy	90% Sunny	90% Rainy	90% Sunny	90% Rainy	0.66

The best score is zero, and it goes to Archangel. The worst is 2, and it goes to Demon. The interesting point is the score differentiates between Cautious and Bold, although both make the same predictions. The better score of Cautious compared to Bold implies that someone who predicts with 60% accuracy should be only 60% confident in her predictions. Bold gets a worse score for being over-confident. The confidence of a prediction is not necessarily a subjective perception in the eyes of the forecaster. The probability that a probabilistic natural law assigns to an outcome can replace the confidence a forecaster puts in her forecast. In this way, the Brier score can rank predictions of alternative probabilistic laws.

One might come up with other scores to measure the success of predictions of a law or a forecaster. For example, the formulation of the Brier score that I presented is suitable only for predictions of binary outcomes such as rain, or no rain. But when the outcome can have a range of values, the above formulation of the Brier score would not be sufficient. We would need an

alternative score that incorporates the precision of the prediction as well. For example, if the temperature is 14 Celsius, a prediction of 14.5 Celsius must get a better score than a prediction of 16 Celsius. The new score should somehow combine accuracy, certainty, and precision. Or one might come up with another score that looks also at how far in the future the forecaster can go. A forecaster who correctly predicts the weather a year from now should get a better score than a forecaster who predicts only tomorrow's weather. Or another score might incorporate the variety of climate types that a forecaster can predict. A forecaster who successfully predicts rainy days only for tropical rain forests would get a lower score compared to a forecaster who predicts rainy days in various geographic regions. The critical point is that no matter what criteria we add, all these alternative scores will put the success of a prediction on a spectrum.

Even the most successful scientific predictions fall short of the absolute ends of success spectrum for various reasons. For example, most of scientific predictions are never maximally certain. The reason is evident if the laws used for prediction are probabilistic. A probabilistic law predicts the outcome with a less than one probability and leaves the possibility that the predicted outcome does not happen. Some laws of science, such as some quantum mechanical laws are explicitly probabilistic. But even some of the laws that look deterministic on their surface are in one way or another probabilistic at their core. For example, chemistry teaches us that carbon dioxide interacts with water and produces carbonic acid. But it is more accurate to say that carbon dioxide interacts with water and produces carbonic acid only if the molecules collide at the correct angle and with sufficient energy. And such a collision is a probabilistic event. Many laws of the special sciences are like this. They are stated with a deterministic tone, but a closer look shows that they are probabilistic at their core.

Scientific predictions are also not maximally accurate and precise because of the noise inherent in any measurement and experiment. Part of this noise comes from the random fluctuations that are always there in the context of any experiment or measurement. And another part comes from the upper cap of precision and accuracy of the measurement devices. But there is also a third source of noise that comes from the limited precision of the values stored in computers. Regular computers have between 7 to 16 decimal digits of precision. Going to more accurate computers mitigates this type of noise but does not reduce it to zero and there is a hard upper cap on the precision of even the ultimate hypothetical super-computer (Davies, 2004).

Scientific predictions also cannot look arbitrarily far into the future. For all the reasons discussed, scientific predicting models all have small systematic or non-systematic errors and those cause problems for predictor simulations. Given long enough, small systematic deviations will

result in humongous differences between the predictions and the actual outcome. In very long time periods, even non-systematic errors might find the chance to accrue in one direction at some point in time, and significantly change the course of the outcome from then on. And both systematic and non-systematic small deviations might significantly deviate the prediction from the actual outcome due to the butterfly effect. Different cases of prediction would be more or less robust over long periods of time and would successfully see further or closer into the future. But no real-world prediction goes until infinite time.

Save the not-yet-found theory of everything, scientific models are also not supposed to predict everything, everywhere. All scientific models work within a limited scope and under certain assumptions. The wider the scope of a model and the fewer its assumptions, the better the model. But the scope never becomes all-encompassing and the assumptions never go to zero. Of course, some level of generality is duly expected from a scientific model. But the generality need not expand to every phenomenon and every situation.

In short, even the best predictions in science are never maximally accurate, precise, and certain, and they are not expected to predict until the infinite future. They are also not expected to predict everything, everywhere. These less than maximal standards apply even to the astonishingly accurate astrophysical predictions and the profoundly accurate scientific devices such as atomic clocks. Maximal standards are too much to ask even from the paradigm successful predictions.

We saw computational irreducibility guarantees impossibility of prediction only for *all cases*, in *infinite time*, or with *infinite precision, accuracy and certainty*. The above discussions, however, show that lack of prediction under these maximal conditions is not equivalent to being unpredictable because maximal standards are not necessary for predictability. It follows that a theory like computational irreducibility that shows some phenomenon cannot be predicted with maximal standards does not show that the phenomenon is necessarily unpredictable.

Computational irreducibility, therefore, cannot provide a basis for unpredictability. We see the philosophical importance of this conclusion in more details in the next section in the context of computational accounts of emergence.

1.4 ONTOLOGICAL WEAK EMERGENCE AND COMPUTATIONAL IRREDUCIBILITY

An emergent phenomenon is a feature of a system in its entirety that looks unpredictable and novel given all that we know about the system's parts. We are confident that we know the micro-structure

of a system well enough to explain and predict its systemic macro properties and behaviours. Yet, the system shows some macro-level features that we cannot predict or explain. There are different ways to explain this dichotomy between what we expect from the micro-structure and what we observe at the macro-level. One way is the approach suggested by Darley (1994) and Bedau (1997, 2008) and supported and expanded by Huneman (2008, 2012). I call this approach *ontological weak emergence*.[7] Due to the heavy reliance of this approach on computational concepts, it is also called *computational* emergence (Huneman, 2008).

Ontological weak emergence sits somewhere between *strong* and *weak* emergence. Strong emergence suggests that emergent phenomena are ontologically distinct from their underlying micro-structure. The most famous examples are mental phenomena that, according to strong emergentists, are of a different nature from the neurological system underneath them (Chalmers, 1996, 2008). Strong emergence is also called *metaphysical emergence*, a term that underlines the ontological distinctness of the emergent phenomena with respect to their underlying micro-structure. Weak emergence, on the other hand, suggests that there is nothing ontologically special about emergent phenomena, and they look novel and unexpected simply due to our limited understanding of the underlying system. In this view, emergence is merely an epistemic perspective in the eyes of the beholder, and it will eventually dissolve as the beholder develops a better understanding of the system (Chalmers, 2008; Hempel & Oppenheim, 1948).[8]

Weak emergence has a special appeal due to the current popularity of physicalism, as it does not introduce any new ontological types and keeps the physicalist ontology clean and tidy. Weak emergence, however, does not explain the notorious persistence and prevalence of emergent phenomena in so many disciplines, from biology to sociology. The history of science shows that many emergent phenomena, such as almost any phenomenon of interest in biology, economics, and so on still count as emergent despite significant advances in more fundamental sciences. The early supporters of weak emergence, such as Hempel and Oppenheim, had a different vision of the future. According to Hempel and Oppenheim (1948), during the course of scientific progress, phenomena are only transiently emergent because the theories and facts to explain them are yet to be discovered. Once the advances in science provide the necessary theories and facts, these will shed light on the phenomenon and dispel the apparent magic of emergence. Although there would

[7] Bedau calls his theory *weak emergence*. I, however, call it *ontological* weak emergence to differentiate it from other accounts of weak emergence.

[8] There are inconsistencies in the field about what is called "weak" or "strong" emergence. I define the terms in the text to avoid any confusion.

be times when "Nature and Nature's laws lay hid in night", the story has a happy ending: "God said, 'Let Newton be!' and all was light."[9]

This has not happened, and many important phenomena in special sciences are still unexplainable and unpredictable by the more fundamental theories (M. Mitchell, 2009). Emergence in special sciences has proved to be way more resilient than what early weak emergentists suggested. This resilience indeed begs an explanation. As Fodor (1997:160-161) nicely puts it:

> Damn near everything we know about the world suggests that unimaginably complicated to-ings and fro-ings of bits and pieces at the extreme microlevel manage somehow to converge on stable macro-level properties … On the other hand, the 'somehow' really is entirely mysterious … So, then, why is there anything except physics? … Well, I admit that I don't know why. I don't even know how to think about why.

This resilience is a good motivation to ask if there is more to emergence than merely the transient surprise of the ignorant. In fact, betraying Hempel and Oppenheim's expectation, not only did the advances in science not demystify emergent phenomena but it was those very advances that led to a stronger ontological version of weak emergence. Studies of complex systems through the 20th century and through today inspired Darley (1994) and Bedau (1997, 2008) to suggest a new kind of weak emergence, namely ontological weak emergence, that while it does not introduce new ontological types, it blames the unpredictability of emergent phenomena on some ontological characteristics of the system. That ontological character is computational irreducibility.

Ontological weak emergence is a property or a behaviour of a system that cannot be explained, derived, or predicted except via simulation (Bedau, 1997, 2008; Darley, 1994). Bedau defines simulation as derivation of the macro-state of the system by going through the course of interactions in its micro-dynamics (Bedau, 1997). The course of interactions can be followed in the system itself, its physical replica, or a computational simulation. The important point is that the only way to derive what comes out of the course of interactions is going through that very course of interactions in one setup or another.

Note that it is not a necessary fact that we need to go through the micro-level interactions to derive the macro-level facts of a system. For example, consider the oscillation of an ideal pendulum. To derive the position of the pendulum at time t ($x(t)$), we do not need to crunch

[9] Epitaph by Alexander Pope

through all the forces acting on the bob in every swing until t. We can simply plug the value for t in the following formula and get the position of the pendulum $x(t)$ (A is amplitude, and ω is angular frequency).

$$x(t) = A\cos(\omega t)$$

But such shortcuts are not available for the ontologically weakly emergent, and the only way to derive them is to go through all the forces step by step. Compare predicting the position of a ball in a pinball machine with predicting the position of a bob of a pendulum. Apparently, we know everything about the mechanics of pinball machines. But unlike the case of the pendulum, we cannot come up with a simple formula to tell the position of the ball at t. The only way to know the position of the ball is to play or simulate each round of pinball. Therefore, the position of a ball in a pinball machine at time t is ontologically weakly emergent.

But why is it the case that simulation is the only way to predict the position of the ball in a pinball machine? Arguably, before the invention of trigonometry by Hipparchus (180 – 125 BC), the shortcut formula for pendulums was not available and simulation was the only way to derive the position of the bob. So, what prevents us from hoping that some future genius will come up with a formula for predicting the position of the ball in a pinball machine, even if it takes devising a whole new branch of mathematics?

In response to this question, Bedau takes recourse to computational irreducibility. He suggests that a phenomenon is ontologically weakly emergent if the path leading to that phenomenon is computationally irreducible. And computational irreducibility guarantees that the only way to predict the phenomenon is going through the original path itself. No shortcuts are possible. In fact, there is such a tight connection between computational irreducibility and ontological weak emergence that Bedau sometimes seems to take the two concepts to be identical (Bedau 2002:18): "The behaviour of [ontologically] weakly emergent systems cannot be determined by any computation that is essentially simpler than the intrinsic natural process by which the system's behaviour is generated. Wolfram … terms these systems 'computationally irreducible'."

Computational irreducibility is an ontological character of the system and therefore, an emergence arising from computational irreducibility is tied to the ontology of the system. Thus, it is not merely in the eyes of some beholder and is here to remain in face of any future advances. Computational irreducibility guarantees that not just us, the mortal humans of the 21st century, but no pinball genius in the future, and no pinball Laplacian demon can tell the position of the ball

without simulating it. Forever and everywhere, simulation is the only option. Or so the supporters of ontological weak emergence claim.

At the core of Bedau's account sits the assumption that computational irreducibility means guaranteed observer-independent unpredictability. The discussions of the previous sections, however, show that computational irreducibility cannot guarantee unpredictability. Computational irreducibility guarantees impossibility of prediction only with maximal standards, and we saw that impossibility of prediction with maximal standards is not equivalent to unpredictability. Computational irreducibility may be able to demonstrate that one cannot predict the position of the ball in a pinball machine under all scenarios, after infinite time, and with infinite accuracy, precision, and certainty, unless one simulates the play. But computational irreducibility has nothing to say against the possibility of a pinball genius, let alone a pinball demon, predicting the position of the ball with strikingly high, yet sub-maximal standards. Computational irreducibility is consistent with the position of the ball being predictable for all that matters.

Avoiding this objection may be the motivation that makes Bedau emphasize *accuracy* and *completeness* in his relatively more recent formulation of ontological weak emergence, saying that it is the accurate and complete derivation of an ontologically weakly emergent phenomenon that is impossible except by simulation (Bedau 2008). But the emphasis on accurate and complete derivation addresses the objection only by trivializing the concept of unpredictability, and consequently the concept of emergence. There is not even a single case of real-world prediction in which one predicts the state of a system with perfect completeness and accuracy. It trivially holds that the only way to know all future states of any physical system with one hundred per cent accuracy and completeness is to run the system itself. But by defining emergence in this way, any physical system whatsoever would turn out to be ontologically weakly emergent. Such an all-inclusive theory of emergence tells hardly anything interesting about the emergent phenomena.

Huneman adopts Bedau's idea but goes one step further. He suggests that on top of unpredictability except by simulation, the non-trivial cases of ontological weak emergence should satisfy one additional criterion. They should show stable higher-level regularities. He again refers to computational irreducibility to explain how an unpredictable micro-level results in a predictable macro-level regularity. He claims that the micro-level is computationally irreducible and hence, unpredictable, but the computational irreducibility is lost once we go to the higher-level (Huneman, 2008, 2012). For example, we saw that coarse-graining transforms a computationally irreducible system to a computationally reducible one. Through a mechanism like coarse-graining,

so the explanation goes, we transit from the computationally irreducible micro-level to the computationally reducible and hence, predictable macro-level.

Similar to Bedau's, a key assumption in Huneman's account is that computational irreducibility means guaranteed observer-independent unpredictability. It is computational irreducibility that supposedly guarantees that emergent phenomena are unpredictable at the micro-level. But we saw that this assumption does not hold, and computational irreducibility cannot guarantee Huneman's criteria of unpredictability on the lower level. His account faces the same problems as Bedau's on that level.

An alternative interpretation of Huneman's account in which the emphasis shifts from predictability on the lower or the higher levels to predictability with maximal or sub-maximal standards seems to fare better. On this interpretation, emergence is unpredictability with maximal standards because of computational irreducibility, while being predictable with sub-maximal standards via processes such as coarse graining that break computational irreducibility. Note that the only role that computational irreducibility plays in this version of Huneman's account is guaranteeing unpredictability with maximal standards, and it is a role that it indeed can play. This guaranteed unpredictability with maximal standards, however, is useless in delineating emergent phenomena in practice, because any natural phenomena you name is practically unpredictable with maximal standards.

This interpretation can at most show a purely theoretical difference between the emergent and the non-emergent, and allows the two classes of phenomena to look exactly the same in the real world. The view, therefore, cannot explain any observable difference between the emergent and the non-emergent phenomena. And the other way round, the practical and observable differences between the emergent and the non-emergent phenomena cannot provide any evidence for this purely theoretical view. The emergence debate, however, is motivated by the supposedly observable and practical differences between the emergent and the non-emergent. With regards to these most eye-catching puzzles of emergence, this purely theoretical approach seems irrelevant and inadequate.

On a third interpretation of Huneman's account, we can focus on predictability after coarse graining rather than unpredictability before it. On this interpretation, we acknowledge that all natural phenomena are unpredictable with maximal standards, and we delineate the emergent as those special ones that can be made predictable by coarse graining. The immediate challenge would be to set a cut off for the amount of coarse graining allowed before an emergent phenomenon becomes predictable. To avoid this difficult challenge, we can define emergence as a graded

concept, suggesting that different phenomena show different amounts of emergence depending on the amount of coarse graining needed before they become predictable. It is an intuitively appealing idea, and I think it is a step in the correct direction as it introduces a graded emergence that reflects the graded nature of predictability. The problem, however, is that it is very hard, maybe impossible, to formalize it as a form of computational emergence. In the next section, I entertain the idea of graded emergence and show that the theory behind computational emergence is too stiff to accompany the moves toward graded emergence.

1.5 DEGREES OF EMERGENCE

One way to deal with the problems arising from the graded nature of predictability is to describe emergence as a graded concept. The promoters of ontological weak emergence have already proposed this idea, though mostly passingly. For example, in his more recent works Bedau (2008) suggests that ontological weak emergence comes in degrees and that phenomena can be more or less emergent. This is a move in the correct direction as it mirrors the graded nature of prediction. The problem, however, is that computational irreducibility cannot accompany emergence in this move. Computational irreducibility works only in extreme and maximal conditions, and it is silent about what falls in the middle. By claiming a graded emergence, Bedau loses the ontological support of computational irreducibility for his theory. This is a heavy loss. After losing the ontological support of computational irreducibility, it is not clear what the advantage of computational emergence would be over the other humbler non-ontological versions of weak emergence. In fact, other more subjective accounts of weak emergence are probably in a better position to define a graded emergence based on grades of predictability, because the spectrum of prediction depends not only on the characteristics of the object of prediction, but also on the subject's epistemic goals and capacities.

The graded version of Huneman's account faces similar problems. The degrees of emergence in Huneman's account start after coarse graining where we can no longer rely on computational irreducibility. We need to devise other computational concepts to formalize and quantify the degree to which a phenomenon becomes predictable after coarse graining. Without such formalization and quantification, the graded version of Huneman's account would not have any ontological and computational element and hence, would not have any tangible advantage over non-ontological types of weak emergence. The non-ontological types of weak emergence might be even better suited to describe the degrees to which a phenomenon shows emergence after coarse graining because they can more easily accommodate the subjective aspects of predictability.

Is there a way to formalize and quantify degrees of emergence in Bedau's or Huneman's accounts? Hovda (2008) has attempted to build such formalization and quantification. His attempt, however, more than anything shows the seemingly insurmountable challenges facing such formalization and quantification. Hovda adopts Bedau's notion of emergence as derivability only via simulation (s-derivability) and suggests the degree to which a phenomenon is emergent corresponds to the amount of simulation, or more precisely, the amount of computation needed to derive that phenomenon. The more computation is needed, the more is the phenomenon emergent.

Despite its apparent simplicity, there are serious problems when it comes to the details of Hovda's formalism. The most important problem that Hovda himself explicitly acknowledges is that the amount of computation needed to derive a phenomenon depends not just on the nature of that phenomenon, but also on somewhat arbitrary choices such as the formal language used to describe the system and the way one defines and counts the steps of the simulation. Listing various such difficulties, Hovda (2008:470) concludes that "we must acknowledge that s-derivability, and amount of simulation required, might be relative to a derivation system" . But this means admitting that the amount of emergence a phenomenon shows at least in part depends on our somewhat arbitrary and subjective choices on the derivation system. The amount of emergence would be a side effect of some arbitrary formalism rather than an ontological fact about the phenomenon.

The dependence of Hovda's quantity of emergence on the choice of formalism diminishes not just its theoretical value in describing the nature of emergence, but also its practical usefulness as a measure of emergence. Science uses different formal frameworks to describe different natural phenomena and therefore, it would not be possible to use Hovda's quantity to compare the amount of emergence across different natural phenomena.

The lessons of Hovda's example are not peculiar to his particular formalism. His work shows how arbitrary and subjective factors bedevil any quantification of weak emergence. Considering these arbitrary and subjective factors, it seems impossibly challenging to quantify emergence as an ontological computational character of the emergent phenomena.

1.6 CONCLUSIONS

Our exploration of various interpretations of computational irreducibility showed that it can guarantee impossibility of prediction only for all cases, or in infinite time, or with infinite precision, accuracy, and certainty. Prediction in its best scientific sense, however, does not happen in these maximal contexts. Every predicting model scores better or worse in terms of the variety of

scenarios it can predict, the stretch of time it can look into, and the accuracy, precision, and certainty of its predictions. Even the archetypes of scientific predictability do not score perfectly on any of these dimensions. Yet, we deem those scientific predictions *successful*, and accordingly, deem the object of their predictions, *predictable*. Therefore, we cannot conclude unpredictability of a phenomenon from a theory like computational irreducibility that shows impossibility of prediction only with maximal standards.

This has important consequences for the philosophical positions such as ontological weak emergence that rely on computational irreducibility to guarantee unpredictability. Ontological weak emergence defines emergence as unpredictability except via simulation. It is computational irreducibility that is supposed to guarantee this unpredictability. Allegedly, computational irreducibility provides an objective and ontological anchor for the unpredictability of the emergent and thus, frees weak emergence from being merely in the eyes of some observer. With its mathematical power, computational irreducibility acts as the seal of Solomon forcing even the Laplacian demons to kneel before the unpredictability of the emergent. Even the demons cannot predict the outcome unless they simulate.

But the discussions above show that until we forge a stronger formulation, the seal will remain lost in the sea.[10] Computational irreducibility leaves open the possibility that not only the Laplacian demons, but even the humble human observers predict the result of a computationally irreducible process with sub-maximal, yet acceptable standards. Computational irreducibility only shows that we cannot ask for the moon. But we do not need the moon after all.

[10] The story goes that Solomon possessed a seal ring by means of which he had dominion over all demons. A demon once stole the ring and threw it into the sea. Solomon was thus deprived of his dominion until he found the ring inside a fish (Jacobs & Seligsohn, 1906).

Chapter 2
Irrational methods suggest indecomposability and emergence

ABSTRACT

This chapter 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.

2.1 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; Kaiser, 2017; Mazzocchi, 2008; M. Mitchell, 2009). 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 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 chapter aims to weaken this counterargument and provide support for alternative non-reductionist metaphysical views grouped under *metaphysical emergence* (Wilson, 2021). Along the way, the chapter 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, 2016d; 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 chapter (Section 2.6) provides more details.

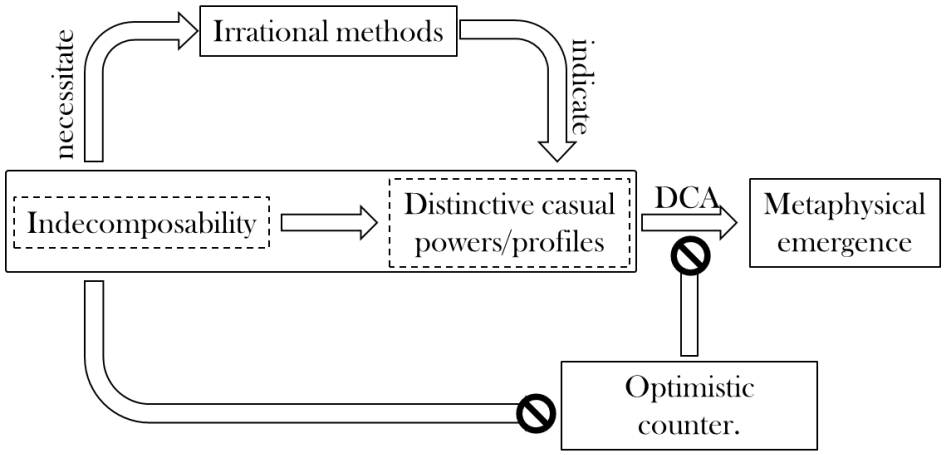


Figure 2.1 Overview of the arguments

I start, in section 2.2, 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 2.3 to 2.6 is schematically shown in Figure 2.1. In section 2.3, 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 2.4, I explain how DCA is blocked by the optimistic counterargument and discuss different versions of this blockade. In section 2.5, 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 2.6, 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.*

2.2 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.[11] 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

[11] For instance, compare the concepts of weak and strong emergence in (Chalmers, 2008; Tabatabaei Ghomi, 2022; Wilson, 2015). 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.

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 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, 2016d; 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 chapter 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.

2.3 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, 2016d; 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.[12] 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 understanding of the fundamentals. This is where the arguments of this chapter come to the aid of the emergentist.

2.4 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. (Batterman, 2001, 2005; Fodor, 1974, 1997; Kaiser, 2017; Mazzocchi, 2008; M. Mitchell, 2009). 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

[12] 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).

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 (2006:432) 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.

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 wonderous 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-19th and early 20th 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.

2.5 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.

2.5.1 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; J. A. Shapiro, 2011; Walsh, 2015b). 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.

2.5.2 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 chapter 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 indecomposibility is indeed the best explanation.

2.5.3 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 Leiber, 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 λ that made its host cells toggle between two gene expression states (Gardner et al., 2000; Khalil & Collins, 2010). Elowitz and Leiber 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 λ 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 aforethought 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 2.6.

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 compeers. 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.

2.6 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.[13] 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.

[13] Bickle, for instance, whom I cited as a prime reductionist, highlights the importance of these methods (Bickle, 2019).

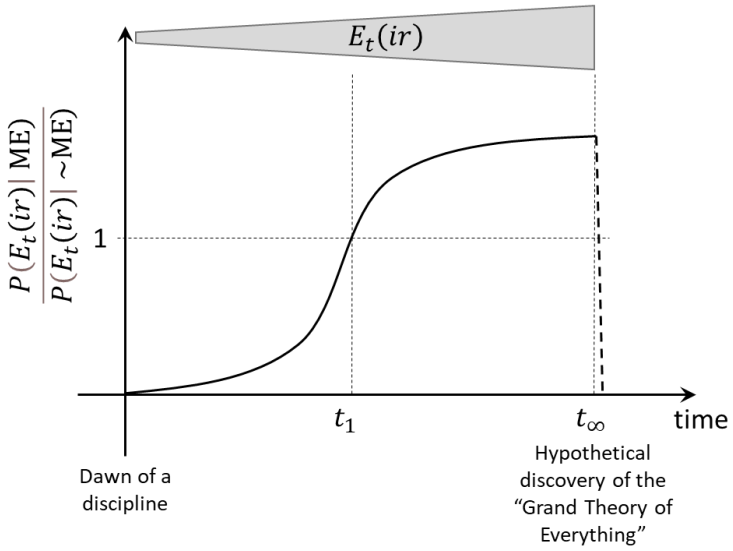


Figure 2.2 Confirmation of metaphysical emergence by evidence of irrationality over time

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 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 Figure 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:

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

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 (~ME). Thus, the above likelihood ratio positively correlates with the following likelihood ratio:

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

Figure 2.2 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 ~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.

2.7 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.*

Chapter 3
A philosophical analysis of the emergence of language

ABSTRACT

There is a research programme in linguistics that is founded on describing language as an emergent phenomenon. This paper clarifies how the core concept of emergence is deployed in this emergentist programme. We show that if one adopts the weak understandings of the concept of language emergence, the emergentist programme is not fundamentally different from the other non-emergentist research programmes in linguistics. On the other hand, if one adopts the stronger understandings of emergence then the programme would have a unique character, but at the cost of some corollaries (philosophical, but not only) which the emergentist linguists would seemingly want to avoid. We show that if the emergentists accept those corollaries, the resulting hypothetical emergentist programme would be totally different from the emergentist programme in its present shape. We conclude that the emergentist programme, as it stands, should be either abandoned, or reshaped in both theory and methodology.

3.1 INTRODUCTION

Emergence has been taken to refer to the apparent irreducibility, or the putative unexplainability and unpredictability of a whole with respect to its constituent parts. This concept has attracted philosophers and scientists alike. There is a large body of philosophical literature on emergence (Humphreys, 2016b; Wilson, 2021), and there are scientists within various disciplines such as biology (Rothschild, 2006; Van Regenmortel, 2004), computer science (Holland, 1998; M. Mitchell, 2009), physics (Anderson, 1972; Wayne & Arciszewski, 2009), and linguistics (Deacon, 2005; MacWhinney, 2002) who have promoted using the concept in their fields. Within the sciences, the employment of the concept is usually sporadic rather than systematic. Linguistics is, relatively speaking, an exception. There is an emergentist research programme in linguistics (from now on, *the emergentist programme*) that is founded on describing language as an emergent phenomenon (O'Grady, 2008), and which aims to explain how humans, of all species, have acquired linguistic

abilities, and particularly, how the child acquires the general rules that govern those abilities. In this paper, we analyse this programme through a philosophical lens.

Our philosophical analysis is aimed to clarify how the concept of emergence is deployed in the emergentist programme in linguistics. As our survey of the emergentist programme shows, linguists understand emergence in different ways which vary in their philosophical content. Some understand emergence in more technical way, and some understand it more philosophically. In each case, we discuss weak and strong understandings of the concept and show that the programme faces a dilemma. On the one hand, if one adopts the weak understandings of the concept of emergence, the concept (philosophically understood) would not add anything significant and the emergentist programme would not be fundamentally different from other research programmes in neuroscience or biology of language more generally (aka *biolinguistics*), and so the there is no need to distinguish the programme by reference to emergence.

On the other hand, the stronger understandings of emergence give the emergentist programme a unique character, but at the cost of some potential philosophical corollaries which the emergentist linguists would seemingly want to avoid. Moreover, we show that an emergentist approach based on the stronger interpretations of emergentism should limit efforts to explain language in terms of factors external to language (biological, social, and the like). Therefore, such a hypothetical programme would not only be very different from the emergentist programme in its present shape, but would go in a direction that is opposite to current biolinguistics approaches, that aim to study language from an interdisciplinary approach and that support the view that cognitive, behavioural, social and cultural facts interact to explain the properties of language, as well as how language is acquired and evolved. Because of this dilemma, we will conclude that the emergentist programme stricto sensu should be abandoned, unless its proponents are ready to adopt a metaphysics (and a research direction) that is incompatible with the current scientific approach to language.

The emergentist linguists who understands emergence as a purely technical, non-philosophical term within linguistics might disagree with our approach and hence, our conclusions. Why, they might think, linguists should care about the philosophical understanding of the term to begin with. However, as we will see through the following survey of the programme, linguists use the term emergence in different ways and not all of those applications are purely technical. As much as they lean towards the philosophical concept of the term, linguists should be aware of the consequences of their philosophically loaded conceptualisations of the term. Even those linguists who use the term in a rather technical way, sometimes make philosophically loaded remarks. A good example

is O'Grady who (as we will discuss in §3.2) has provided a technical description of the emergentist programme and yet, begins his discussions by defining emergence as "the process whereby the interaction of simple entities, forces and events produces a system with its own novel properties." (O'Grady, 2021:1). We will show in section 4 that emergentist claims of novelty and uniqueness have important philosophical ramifications. Eventually, the conclusion of our discussions for an emergentist linguist who insists on a strictly technical, non-philosophical understanding of the term could be that it is better that linguists abandon using such a philosophically loaded term altogether, and use other more benign terms to avoid potential confusions and misconceptions.

On the other hand, philosophers of emergence will benefit from our linguistic discussions by seeing philosophy of emergence in action. Broadly speaking, the puzzles of language emergence are not fundamentally different from the general puzzles of emergence, and the emergentist programme within linguistics, as a relatively speaking well-developed scientific emergentist programme, provides a particularly interesting case study for philosophers. And the other way round, the case also provides an opportunity for philosophical insight to contribute to science by explicating the key foundational concept of a scientific programme.

We start with some preliminary discussions of the philosophical concept of emergence in section 2, and characterise the concept as simultaneous dependence and independence of a whole with respect to its parts. We then switch to linguistics and give an overview of the emergentist programme in section 3. In section 4, we hypothesize about a truly emergentist programme in light of the philosophical concept of emergence, as discussed in section 2. We discuss weak and strong views of independence, and analyse the philosophical and methodological consequences of each of these understandings for language. The analyses of some of the strong understandings lead us to discuss Kim's exclusion argument, one of the most important challenges facing emergentism, in section 5. Finally, in section 6, we present some conclusions about our philosophical analysis of the concept of emergence as applied to language. We suggest that the core lesson drawn from our analysis can inform similar applications of the concept in other disciplines such as biology where the concept is gaining popularity (Mazzocchi, 2008; Newman et al., 2003; Newman & Comper, 1990; Oyama, 2000; Van Regenmortel, 2004). At the end of the paper, we summarise this generalisable core conclusion in the following adage: Half-hearted emergentism is expendable. Wholehearted emergentism is expensive.

3.2 EMERGENCE: SOME PRELIMINARY CONSIDERATIONS

Emergence is usually associated with the old slogan that the whole is more than the sum of the parts. Philosophers of emergence have explicated this slogan in different ways. But most of those explications, one way or another, describe emergence as some high-level properties of hierarchical systems showing some sort of independence with respect to the lower level constituents they also depend on, in the sense that their properties cannot be reduced to, or explained and predicted by the properties of the constituents at lower levels of the hierarchy (Humphreys, 2016a). This apparently paradoxical dependent and independent nature of emergent phenomena is the distinctive feature of emergence.

The independence is usually demonstrated by referring to features that are idiosyncratic, that is, distinctive with respect to low-level features. Consider the case of mental phenomena, like feeling pain. The sense of pain depends on the underlying neurobiological machinery, specifically, the nociception system, involving skin receptors, specific brain areas, and the like. But at the same time, feeling pain also involves an introspective phenomenological perception of pain, which can be argued to be of a different nature compared to its underlying neurological machineries. Accordingly, one could argue that a hypothetical neurologist who knows everything about nociception but has never experienced pain, does not really know what pain truly is (Chalmers, 1996; Nagel, 1974). Importantly too, a wildly wide range of neurobiological machineries and states can all result in a perception that is nonetheless called "pain" (i.e. the multiple realisability of pain), so that pain cannot be exactly equated with any of those underlying machineries, but with the sum of all of them (Aizawa, 2013; Fodor, 1974). Therefore, at least on the face of it, it seems that pain both depends on, yet it is independent of, the underlying neurobiological machinery. As noted, this paradox is the essence of emergence and so we can conclude that pain is an emergent phenomenon.

Different theories of emergence diverge by the various ways that they describe and explain this simultaneous dependence and independence of emergent phenomena. These theories fall into two broad categories, metaphysical, and epistemic (Chalmers, 2008; Wilson, 2021). Theories of metaphysical emergence suggest that emergent phenomena, once produced, are metaphysically distinct from their lower-level constituents, this accounting for their distinctive properties. If pain is metaphysically distinct from the underlying nociception system, it is not surprising that it shows its own unique features. Theories of epistemic emergence, on the other hand, deny metaphysical independence of emergent phenomena and suggest instead that their apparent independence with respect to their lower-level components is only due to epistemic limitations of beholders such as

us, humans. From e.g. God's perspective, so the thought goes, pain is nothing but whatever its underlying producing systems are. But for us, because of our limited knowledge, pain cannot be explained by reference to those underlying systems only, and this is why it appears to have some unique properties.

Epistemic approaches to emergence divide on whether they take the epistemic limitations that result in appearance of emergence to be surmountable, or insurmountable. Those that take the epistemic limitations to be surmountable, suggest that emergence is a passing side effect of our current technical, and perhaps theoretical limitations, and therefore, they predict that it is going to fade away in light of future technological and scientific advances (Hempel & Oppenheim, 1948). On the other hand, those that take the epistemic limitations to be insurmountable contend that emergence will forever be part of our scientific understanding of the world (Bedau, 1997).

The concept of emergence has been deployed in a variety of scientific contexts, though the scientific value of the concept has been a matter of debate. Modern biology, for example, is a field that abounds with emergentist claims (Heng, 2017; Kaiser, 2017; Kauffman, 1993; Mazzocchi, 2008, 2011; Mikulecky, 2001; Plsek & Greenhalgh, 2001; Rickles et al., 2007; J. A. Shapiro, 2011; Tabatabaei Ghomi, 2023; Walsh, 2015b). Observations such as asymmetries in relative prevalence of sugar or amino acid enantiomers (Anderson, 1972), concepts such as 'organism' (Walsh, 2015b), or phenomena such as biological programming (Noble, 2008) are claimed to be best explained within an emergentist framework. Importantly, scientific recourses to emergence are usually not merely theoretical viewpoints, but also have a methodological dimension. If one takes emergent phenomena to be metaphysically irreducible to their constituent parts, or epistemically unexplainable in terms of interactions between parts, then it may make no sense to try to understand those phenomena by reference to such low-level mechanisms. Instead, one should better consider the system as a whole and try to describe and explain its features on the systemic level. In biology, for example, some emergentists claim that for understanding what an organism is, instead of zooming in and merely describing how its components work, we should better zoom out and describe its characteristics from a holistic perspective (Tabatabaei Ghomi, 2023; Walsh, 2015b). Similar views motivate practical fields such as systems pharmacology that promote finding medical solutions not by modulating the effects of a particular drug target, but by intervening with the dynamics of molecular and physiological networks as a whole (Latourelle et al., 2017; Luu & Palczewski, 2018; Wang et al., 2021).

Our focus in this paper is on language. We aim to understand in which sense, if any, language can be regarded an emergent phenomenon, and to explore the ensuing philosophical and

methodological consequences. We will ask how the approaches to language that are regarded as *emergentist* characterise the emergence of language, and we then discuss the different philosophical ways to understand each of those characterisations. In this, we will examine more closely the two types of theories of emergence mentioned above, namely, metaphysical and epistemic, as applied to language. In every case we discuss the potentially problematic aspects resulting from each understanding of language emergence, possible solutions to address those complications and ultimately, the kind of emergentist programme that ensues from those solutions. Eventually, we will reveal the dilemma facing the emergentist programme: The weak understandings of language emergence do not warrant establishing a standalone emergentist programme, and the strong understandings of language emergence come with philosophical and methodological consequences that the language emergentists seemingly prefer to avoid.

3.3 A GENERAL OVERVIEW OF EMERGENTIST APPROACHES TO LANGUAGE

Usually, emergentism in linguistics refers to the view promoted by Bates (1998), MacWhinney (2002), O'Grady (2021) and others that the language faculty is a new machinery built out of old parts, as opposed to the "essentialist" or "nativist" view promoted by Chomsky (until recently) and many of his followers (Berwick & Chomsky, 2019) that the language faculty is built on inborn universal grammatical principles that have no counterpart elsewhere in cognition. Despite this prevalent understanding of language emergentism, there are different ways that the term "emergence" is used by different linguists, and it is dubious that a solid, well-founded emergentist programme of language exists. What one finds is a potpourri of sometimes half-baked ideas. As Austin (2021:78) puts it, emergentist formulations are still "brainstorming more than the kind of definition of language that is needed for a scientific linguistics". This situation makes it almost impossible to formulate a unifying definition of *language emergence* or even *emergent properties of language* that accurately applies to all the so-called emergentist approaches to language, and particularly, that is compatible with how *emergence* is construed in philosophy, which is the main target of our approach here. Considering this situation, we try to summarize (and eventually, clarify) the different ways that the concept of language emergence is characterised in the field of linguistics. For each characterization, we consider its plausibility and implications according to extant philosophical approaches to emergence. We also ask if that particular characterisation of language emergence can sufficiently differentiate from non-emergentist accounts. More specifically, in each case we will ask if approaches avoiding the concept of emergence would really provide poorer, less

biologically accurate views of language, as claimed. Ultimately, if non-emergentist and emergentist views of language overlap or prove to be equally accurate and useful, we will discuss if it makes sense to maintain a differentiated emergentist research program in linguistics.

3.3.1 two types of language emergence

Our analysis of the so-called emergentist literature in linguistics suggests that two different views of language emergence exist in this field. First, the emergence of language out of non-linguistic components, both during language acquisition by the child and during language evolution in the species (for example, see Deacon, 2005; O'Grady, 2005). Second, the emergence of language phenomena within the realm of language, such as the semantic enrichment of words when they are used in specific utterances to convey context-dependent meanings. As a simple example, consider the meaning implied by the word "sun" when someone quotes the last verse of Shakespeare's sonnet 33: "Suns of the world may stain when heaven's sun staineth." In this case, the meaning of the word "sun" is said to emerge from the conceptual, grammatical, and pragmatic systems in which it is plugged (see also MacWhinney, 2005). This type of language emergence forms the basis of the connectionist methods of language analyses that try to parse meanings and grammatical structures based on contextual cues (Palmer-Brown et al., 2002). Putting this differently, the first type of emergence is about how language develops, whereas the second one is about how language works once it is developed. The focus in the paper is put on the first type of language emergence.

That said, these two types of emergence are commonly mixed together in discussions of language emergence, so that both types are put under the single banner of language emergentism. This mixing is not due to scholarly sloppiness. These two types of emergence are, in fact, related. How language works depends on how it is acquired, which depends in turn on how it evolved. According to the first type of language emergence, language acquisition entails the development and the progressive functional interrelationship of diverse auditory, visual, cognitive, and behavioural mechanisms that evolved for other functions, but that have been recruited for language during our recent evolution (Kandel & Hawkins, 1992; MacWhinney, 2002). This can be seen particularly well at the neurobiological level, where language processing by the brain depends on neural devices that perform basic representations and computations and that are recruited for other cognitive functions besides language (Poeppel, 2012; Poeppel & Embick, 2005), with this language machinery showing a strong evolutionary continuity with devices found in other animals, particularly, primates (e.g. Lieberman, 2016). It makes sense that the ultimate linguistic products

of such a background can be understood only by concerted consideration of various types of inputs, rules, and contextual dependencies, leading to the second type of language emergence.

Despite their relations, however, these two types of language emergence can be distinguished ontologically, because they can be said to happen, as noted, at two different borders. The first type happens at the border between biology and linguistics. The second type happens within the realm of linguistics. As also noted, our focus will be on the first type, but we will discuss aspects of the second type where they become relevant to the discussions of the first type.

3.3.2 A negative characterization of language emergence

One classical formulation of language emergence (of the first type) is O'Grady's:

> The phenomena of language are best explained by reference to more basic non-linguistic (i.e., 'non-grammatical') factors and their interaction—physiology, perception, processing, working memory, pragmatics, social interaction, properties of the input, the learning mechanisms, and so on. (O'Grady, 2008:448)

On the face of it, one could argue that this formulation just emphasises the dependence of language on non-linguistic factors but fails to clarify the more important aspect of the alleged emergent nature of language, that is how language is independent from such non-linguistic factors. Probably, O'Grady's purportedly emergentist approach was originally meant to separate the functionalist characterization of language development from nativist accounts, particularly, Chomsky's view. As noted at the beginning of §3, it is common to describe the emergentist approaches in contrast with the nativist research programme (MacWhinney, 2002, 2015; O'Grady, 2008). Roughly speaking, the nativist approach sees language as a set of fixed rules resulting from a certain language-specific machinery in human brains that works under some language-specific constraints and that evolved to fulfil some particular functions, specifically, generating more sophisticated thoughts (Berwick & Chomsky, 2016; Chomsky, 2005; Hauser et al., 2002; Pinker, 1994). This view is notoriously in debt with Fodor's (e.g. 1983) view of cognition and the subsequent takes by evolutionary psychology (e.g. Sperber, 2001). Instead, the emergentist programme, as formulated by O'Grady and others, describes language as a network of assorted items, such as sounds, meanings, combinatorial rules, and the like, otherwise not specific to language and resulting from complex interactions between innumerable biological, cognitive, social, environmental, mathematical, and other factors (MacWhinney, 2015). Furthermore, it supports the view that language acquisition depends on perceptual and cognitive abilities neither specific to language, such as categorization, statistical learning, and the like (e.g. Saffran et al.,

1996). Ultimately, it supports the view that if language can be eventually characterized as a distinctive, full-fledged cognitive faculty within the human brain/mind, this is only true in the adult state, so that language is more an ontogenetic outcome than something existing ab initio. This emergentist program when it comes to cognition and the brain can be labelled as *neuroconstructivist* (following Karmiloff-Smith, e.g. 2009, Sirois et al., 2008, and others), because as also famously stated by Bates and colleagues (1988:284), "modules are not born, they are made."

In view of this, it can be argued that the nativist programme and the emergentist programme fundamentally differ when it comes to the second type of language emergence. The nativist programme suggests that language is shaped primarily by rules that are specific to language, while the emergentist programme denies the existence of such specific rules, and instead, suggests that linguistic rules do not differ from other rules governing cognition, so that the rule-like regularities of language emerge out of complex interactions beyond language. The two programmes also differ with respect to the first type of language emergence. These differences were more noticeable in the past, but as we will show below, they have become less and less noticeable as our understanding of how the brain works and how it evolved has improved.

The nativist programme suggests that there is a specialised language machinery in the human brain which grows in the child under genetic guidance mostly, and that results in a set of rules specific to language. This claim was stronger in the first proposals by Chomsky, as with the UG (Universal Grammar) model. As noted by Chomsky (1977:164), UG is "a common human attribute, genetically determined, one component of the human mind". By contrast, the emergentist programme denies that there is such a specialised machinery, so that the language-specific patterns of brain activity that can be identified during language processing result, as noted, from complex interactions between non-linguistic brain devices. That said, there is quite a noticeable common ground between these two programs. On the one hand, it can be argued that this type of recurrent patterns of brain interconnection comprises a module, at least a functional or mental module, in the spirit of current evo-devo approaches in biology (see Breuker et al., 2006, or Griffiths, 2007 among many others). On the other hand, if one considers the last proposals by Chomsky, particularly, after the advent of the Minimalist Program (MP) (2005), language can be viewed as well as arising from biological infrastructures. In the MP, the computational system of language is reduced to a single operation *Merge*, and language ultimately results from an interface between the conceptual-intentional system, important for meaning, and a sensory-motor system involved in the perception and production of linguistic items (see Hauser et al., 2002 for details).

Despite these biological grounds, standard minimalism insists on at least a minimal degree of language-specificity in Merge (Berwick & Chomsky, 2016, 2019) and has it that this specifically linguistic Merge is what differentiates their program from reductionist or "negative" emergentist approaches such as O'Grady's. By claiming some aspects of Merge to be exclusively linguistic, standard minimalism introduces a linguistic element that is irreducible and thus, independent of non-linguistic cognitive elements and therefore, falls under positive accounts of language emergence that are discussed in the following sections. However, many adherents of minimalism have departed from standard minimalism, specifically when discussing language evolution, and have argued that Merge could derive from a non-specifically linguistic ability. For example, Fujita & Fujita (2022) argue that Merge could derive from motor action planning, Liu et al (2023) suggest that Merge can be localized in Broca, as a non-language-specific brain area, and others have suggested that the computational system of language can be decomposed in primitives (a sequencer, a memory stack) that can be found in other animals contributing to other functions (Balari, Antonio, et al., 2013; Balari, Lorenzo, et al., 2013; Lieberman, 2000). Such proposals are in line with a general trend in in neurolinguistic research that suggests language can be spelled out in terms of basic components that are not specific to language, following e.g. Poeppel (2011, 2012). On these non-standard views, the MP can be seen as a variety of negative emergentism described by O'Grady's description of language emergence, that language is best described in terms of its non-linguistic underlying causes.

Another aspect of the MP that is emergentist by O'Grady's formulation of language emergence is the role of what Chomsky calls the "third factor". The third factor refers to non-linguistics principles and constraints that affect the development and eventually, the final shape of language (Chomsky, 2001: 1-2, 2005). For instance, take the mathematical principles of efficient computation. These principles are inherently non-linguistic but influence the computational system of language.[14] Biolinguistics can offer a wide range of such third factor influencers, like various physico-bio-chemical parameters and properties such as viscoelasticity, differential biochemical diffusion and oscillation, mechanochemical excitability, and the very dimensions of the space in which chemical reactions take place. Newman and Comper call such factors "generic factors", and show their roles in biological pattern formation in general (Newman et al., 2003; Newman & Comper, 1990).[15] Patterns of language are no exceptions. Under the influence of these

[14] We have to underline that the minimalist MP is far away from providing a successful detailed third factors explanation for linguistic phenomena.

[15] Similar ideas can be seen in works of Kauffman (1995, 2000) where he writes about the general laws which regulate the self-organization of biological systems, and Oyama's when she claims that part of the information which determines the features and functional properties of any biological structure is generated by developmental processes themselves (Oyama, 2000; Oyama et al., 2001).

third factor principles, many features of language "come for free" from non-linguistic origins, that is, are emergent by nature, thus approaching Chomsky's MP to the allegedly rivalling emergentist programme as described by O'Grady.

In summary, once one appreciates that the differences between the recent views in the nativist and the emergentist programmes are unessential and diminishing, O'Grady's formulation of language emergence loses its significance, and seems to be a relic of outdated debates. The formulation only tells us that there is no biological machinery in the brain operating under language-specific principles and resulting from specific genetic cues. As noted, most neurolinguists, psycholinguists, and biolinguists would now agree on this. More importantly for our philosophical concerns here, the formulation gives us only a negative understanding of the concept of language emergence by rejecting some outdated views about the origins and the development of language, but does not provide a positive understanding of language emergence by clarifying the role that emergence plays in the nature of language, and particularly, in how it is acquired, how it evolved, and how it operates. Going back to our philosophical characterisation of emergence, O'Grady's formulation of language highlights the *dependence* of emergence on non-linguistic elements, and fails to characterize how language is *independent* and distinct with respect to its non-linguistic precursors. In other words, O'Grady seems to be a complete reductionist about language, with his account telling that the whole is best described in terms of its parts, and his account does not follow the old emergentist slogan that the whole is more than the sum of the parts.

This strong reductionist nature of the negative characterization of language emergence has been noticed with concern by some linguists whose sense of language emergence is closer to the philosophical anti-reductionistic sense of the term. Geoff Jordan, for instance, writes:

> I think the latest radical empiricist versions of emergentism are actually quite dangerous; no more dangerous than one of those meteorites that might collide with planet earth, but dangerous. While classic emergentism was unable to explain how novel properties could emerge from complex systems, and thus remained somewhat mysterious (even smacking of dialectics), the latest versions of emergentism seem to be getting closer to a model of the process. The problem with this for a rationalist … is that the more it becomes possible to demonstrate the systematic interconnections between psychology and physics, for example (the more we can do away with the construct of the mind, and just talk about the brain), the closer we get to describing the necessary and sufficient conditions for psychological states in physical terms, and the closer we get to reductionism. Reductionism finds the ultimate meaning

of the "object" not in its inherent qualities but in the parts which compose it, which is to say that we enter the topsy turvey world where there are only parts. (Jordan, 2003:247)

However, there are other language emergentists that have leaned, to different degrees, towards a positive characterization of language emergence with anti-reductionistic features as we discus below.

3.3.3 Towards a positive characterization of language emergence

MacWhinney (2015) has suggested another unifying theme for the emergentist approach to language that gets at least one step closer to a positive characterisation of language emergence. He suggests that emergentism in linguistics means working within three frameworks: emphasis on Darwinian evolution, analysis of complex systems as structured hierarchies, and the appreciation that the processes on each level of these hierarchies happen at different timeframes. MacWhinney claims that these three frameworks are the common themes of all emergentist accounts of language, and the differences between them arise only from which framework they emphasise.

One problem with MacWhinney's thesis is that the two types of language emergence, as characterized in section 3.1 above, are mixed in its various aspects. For instance, the Darwinian evolution that MacWhinney refers to includes both the evolutionary processes resulting in the human language-ready brain, and the evolutionary relations that supposedly exist between linguistic components that compete to serve communicative functions (MacWhinney, 2015). The two types of language emergence are also mixed up in MacWhinney's description of structured hierarchies. MacWhinney (2009) introduces six hierarchical levels in language: auditory phonology, articulatory phonology, lexicon, syntax, embodied roles, and communicative structure. However, whereas some of these levels are within the realm of language, others have non-linguistic nature (e.g. embodiment). Finally, the same mixing happens when MacWhinney describes the different timeframes associated with the different levels of his hierarchy. MacWhinney (2015) suggests four timeframes for various levels of language production: the timeframe of processing that happens at the moment of speaking, the timeframe of consolidation of experiential traces into memory, the timeframe of social diffusion of linguistic forms, and the timeframe of diffusion and consolidation of genetic basis of linguistic abilities.[16] This list is a mixture of biological, social, and linguistic mechanisms. As noted above, the two types of language emergence are related and therefore, it is not necessarily wrong to mix them together. But unless this is explicitly acknowledged and properly characterized, the amount of mixing and generalisation in MacWhinney's characterisation

[16] To be more precise, these are space-time frames and have spatial, as well of temporal aspects.

precludes to inform a precise understanding of the phenomena pertaining to the biological machinery of language vs. the phenomena pertaining to language (and accordingly, to the first or the second types of language emergence).

Besides this issue, the main problem of MacWhinney's proposal is that, similar to O'Grady's thesis, it is completely compatible with a non-emergentist approach to linguistics. One can agree that evolution has played an important role in language development, that the underlying causal mechanisms of language have a hierarchical structure, and that different causal mechanisms work on different timeframes without being obliged to accept that language has an emergent nature in the sense we characterized *emergence* in section 2. In absence of further evidence, there seems to be no need to bring emergence up.

Perhaps one of the most elaborate emergentist proposals of this 'positive' type is Deacon's (2005, 2014), who promotes the view, specifically, that articulated i.e. linguistic thought depends on unarticulated thought. According to Deacon (2005), there are infra-linguistic hierarchies of non-linguistic cognitive, semiotic, and pragmatic units of thought that are precursors to the articulated linguistic components. These precursors are present and active synchronically all along the formation and articulation of linguistic components. As the final products of language articulation are built on this concurrent infrastructure of precursors, they are hierarchical, rather than temporal precursors of language. Actually, Deacon suggests that the non-linguistic infrastructures operate on slower timeframes compared to language articulation. Accordingly, one can utter several sentences while being in the same general mood or have the same general unarticulated thought. In line with current neurolinguistic hypotheses, Deacon associates different brain areas with different elements of these infra-linguistic hierarchies, so that language ultimately depends on several of these functions working together. Finally, he also suggests that while language is an exclusively human phenomenon, many forms of unarticulated thought seem to be common between humans and non-humans such as primates. Again, as with MacWhinney's proposal, it is not clear in which sense Deacon's view benefits from bringing *emergence* to the forefront. However, Deacon has more to say about language emergence, and below we will discuss his other views in more details.

In the next section, we will show that it is possible to support the view that language is emergent by nature. But for this we need to pass on these general characterisations and follow the conceptions of emergence more closely, with the aim of probing not only that language can be dependent on non-linguistic infrastructures, but particularly, that it can be independent from them too (i.e. what we have called a *positive* characterization of language emergence). The

characterizations above only attend to the dependence of language on the underlying infrastructure and non-linguistic precursors. Many other examples can be found in the modern biolinguistic literature, particularly, about how language evolved. One case is the claim that language evolved in connection to other non-linguistic adaptations such as adoption of a bipedal gate, developing control over phonation, increased brain size, the formation of social support for child rearing, and so forth (Benítez-Burraco & Nikolsky, 2023; MacWhinney, 2002). Another example, related to this, is the claim that there are no qualitative systematic differences between the human brain that shows linguistic abilities, and the brain of other primates that lacks any such abilities (Deacon, 2005; Lecours et al., 1983), so that language evolution, like the evolution of many other human-specific abilities, can be mostly viewed as the outcome of the kludge process that put into contact, and perhaps optimized, previously evolved brain mechanisms (see Marcus, 2008), actually, in the line of Chomsky's MP. But these dependence relations are not enough for language to be emergent in the sense characterised in section 2. We also need to probe, as noted, that language has some sort of independence from the non-linguistics precursors and infrastructure.

3.4 CHARACTERISING LANGUAGE EMERGENCE BY INDEPENDENCE RELATIONS

In this section, we characterise language emergence of the first type by looking for different ways that language is described to be independent of its non-linguistic infrastructure in the emergentist literature. We further discuss the various understandings of those relations and their philosophical consequences.

Consider several core units in the structural characterization of language: phonemes (sensorily identifiable units of sound that allow distinguishing words from one another); morphemes (indivisible meaningful linguistic units within words), syntactic rules (instructions about combining words to convey complex meanings), etc. As discussed above, diverse non-linguistic infrastructures underly the representation of these core units and their computation: mathematical principles of information encapsulation, thoughts and moods, cognitive mechanisms of learning, biological sensory mechanisms, etc. We now wish to understand what it means that linguistic components are independent from non-linguistic infrastructures, as independence relations are crucial, as noted, for distinguishing truly emergentist proposals from other characterizations of language that also claim that language develops and evolves from the interaction between different non-linguistic components: perceptual, cognitive, and behavioural, as the proposals discussed in the previous section.

The independence of language becomes manifest in its distinctive features, if any, that are peculiar to language and are absent from the non-linguistic infrastructures. Such distinctive features have been referred to as the hallmarks of emergence even by those linguists who define language emergence in a rather technical way. For instance, we saw at the beginning of the paper that even O'Grady associates emergence with a "system with its own novel properties." (O'Grady, 2021:1) In the emergentist literature within linguistics, one can identify at least four ways that language is described to be distinct and independent with respect to its underlying non-linguistic infrastructures: i) unpredictability by the underlying mechanisms (Bates et al., 1998; Deacon, 2003; MacWhinney, 2015; O'Grady, 2021); ii) unexplainability in terms of lower-level entities (Deacon, 2005; MacWhinney, 2015; McClelland, 1987); iii) possession of new causal powers and performance of distinctive functions (Deacon, 2003); and iv) having unique properties such as an organismic life in a linguistic ecosystem (Croft, 2000; Frank, 2008; Piattelli-Palmarini & Uriagereka, 2004; Ritt, 2004). In what follows we analyse each of these types of putatively distinct characters of language. In each case, we discuss various ways that one can understand these claims of distinctness and independence, and the ensuing philosophical consequences.

3.4.1 Unpredictability

Unpredictability is frequently cited in philosophical discussions as a hallmark of emergence (Bedau, 1997; Huneman, 2008). The idea is that in non-emergent systems, understanding the underlying mechanisms suffices for predicting how the system would behave. For instance, if I know exactly how my car works, I can predict what happens when I turn the steering wheel to the right. In emergent systems, however, even a complete understanding of the underlying mechanisms is claimed to be insufficient for predicting how the system would behave. The weather is an example. Even if we know all the underlying factors contributing to a weather system, so the thought goes, we cannot predict the weather for the next week with one hundred percent certainty. In this line, some emergentist accounts of language have sometimes claimed that even if one identifies and understands all the non-linguistic infrastructures that contribute to language, one still cannot predict the final shape of language that results from those mechanisms (Bates et al., 1998; O'Grady, 2021). O'Grady, for instance, writes:

> A defining feature of complex systems is the presence of emergence; they become more than the sum of their parts, taking on properties and manifesting effects that could not have been predicted in advance. (O'Grady, 2021:7)

Unpredictability is closely tied to unexplainability. So, we discuss its various interpretations and philosophical implications along with unexplainability.

3.4.2 Unexplainability

Another related type of independence that emergentists have attributed to language is its unexplainability by non-linguistic infrastructures. This unexplainability is best understood in terms of the indecomposability of emergent phenomena. Decomposition is the method of understanding a phenomenon or entity by disintegrating it into its constituent parts and the interactions between those parts (Bechtel & Richardson, 2010). It is probably the most prevalent method of explanation for composite entities. In simple terms, decomposition means understanding by reverse-engineering. Accordingly, an entity or phenomenon that is fully decomposable to its constituent parts and their interactions is not emergent. Being fully decomposable means that the entity or phenomenon is basically nothing but the parts and their interactions and, therefore, cannot have any independence and distinctness with respect to those parts and their interactions. On the other hand, indecomposable entities cannot be explained away by decomposing them to their constituent parts and their interactions and, consequently, show some sort of distinctness and independence with respect to their parts and their interactions. This independence makes them emergent.

The philosophical literature on emergence divides on the *in principle* vs *in practice* unpredictability and unexplainability of emergent phenomena. Some theorists ascribe the apparent unpredictability and unexplainability of emergent phenomena to our limited understanding of those phenomena at this present time. For this group, the emergent properties of systems are unpredictable and unexplainable only in practice. They suggest that as our understanding of emergent phenomena progresses over time and we build stronger experimental, computational and modelling tools, we will eventually be able to explain and predict the purportedly emergent phenomena by means of their lower-level infrastructures. At that point in time, emergence would fade away (Hempel & Oppenheim, 1948). For this group of theorists, emergence is merely a theoretical placeholder until we figure things out. This is an anti-emergentist position that explains away the appearance of emergence as a side effect of a passing situation.

Some other theorists, on the other hand, suggest that the difficulties associated with explaining and predicting emergent phenomena are insurmountable in principle. For instance, theorists of computational emergence suggest that the unpredictability of emergent phenomena arises from certain computational characteristics of emergent systems such as their so-called computational irreducibility that make their emergent features ever-unpredictable, either by us, or by any other natural intelligence (Bedau, 1997, 2008; Huneman, 2008). In the domain of physics, some theorists have suggested that the fusion of some particles can result in new (emergent) entities that are not decomposable and thus, not explainable any more by the isolated individual constituents

(Humphreys, 1997). Finally, in the domain of biology, some researchers have proposed that complex systems such as organisms are inherently indecomposable and thus, can be correctly explained only if we go beyond their constituent parts, and acknowledge their irreducible systemic characters (J. A. Shapiro, 2011; Walsh, 2015b).

It is important to understand where theories of language as an emergent phenomenon stand on this debate of in principle vs in practice unpredictability and unexplainability. In general, the reasons that emergentist linguists cite for unpredictability of language are mathematical and point towards an *in principle* unpredictability of the emergent features of language. They attribute the unpredictability of emergent features to the mathematical framework that produces those features. Bates et al, for instance, write:

> In an emergentist theory, outcomes can arise for reasons that are not predictable from any of the individual inputs to the problem. … [I]t has been argued that grammars represent the class of possible solutions to the problem of mapping hyperdimensional meanings onto a low-dimensional channel, heavily constrained by the limits of human information processing (e.g., MacWhinney & Bates, 1989). Logic, knowledge and grammar are not given in the world, but neither are they given in the genes. (Bates et al., 1998:590)

Another example is Deacon (2003) who has referenced the chaotic nature of systems such as non-linguistic infrastructures and the unpredictability that follows from the so-called butterfly effect (Lorenz, 1972, 1993).

That said, unpredictability by itself, even if it is *in principle*, is not enough to capture the significance of language as an emergent phenomenon. A random pattern resulting from a random generator might be unpredictable in principle for the same mathematical reasons. Emergence of language should be something more significant than mere unpredictability. Unexplainability fares better in capturing this significance. The reason that emergentists sometimes bring for unexplainability of language, namely, its indecomposability, makes language unexplainable *in principle*. If language is indecomposable by nature, then by definition, it cannot be explained by decomposing it to its underlying non-linguistic infrastructure. For instance, Deacon (2005) refers to the failure of many attempts to reverse-engineer the formal structure of language to neurological, genetic, or algorithmic infrastructures and takes this as evidence that language is not completely decomposable to lower-level non-linguistic infrastructures. Similar partial indecomposability is also highlighted by MacWhinney (2015).

Paradoxically, despite their claims of indecomposability of language, emergentists constantly try to explain language by decomposing it to non-linguistic infrastructures. MacWhinney (2002), for one example, tries to explain language by reference to various biological and social underlying mechanisms. Such reductionistic approaches are so prevalent in the emergentist approaches to language that O'Grady (2008) sees explaining language in terms of non-linguistic causes to be the unifying theses of these approaches (see his "emergentist thesis" discussed above), and Jordan (2003) finds those approaches "dangerously" reductionist (see the quote towards the end of §3.2). Considering the prevalence of such reductionist explanatory attempts in the emergentist approaches to language, we should give a very weak interpretation to many such emergentists' claims of indecomposability and unexplainability. On a weak interpretation, indecomposability of language could merely mean that language cannot be directly linked to isolated causes, and it results from multiple causes acting together with reciprocal interactions. But this weak interpretation does not make language unexplainable in principle and, accordingly, not emergent either. Arguably, almost any phenomenon of scientific interest results from multiple causes acting in complicated ways and non-emergentist, mechanistic explanations have successfully explained many such phenomena. We need more than this to justify allocating language in a special "emergent" category.

One may underline that scientists need to go beyond traditional mechanistic approaches to explain complex phenomena and suggest that this methodological necessity shows that those phenomena are emergent. A good example is capturing complex biological links in systems biology: contrary to other approaches to biological phenomena, which are methodologically reductionist, systems biology tries to understand and explain the organismal structure and function by focusing on properties of the whole system (Kitano, 2002). Systems biology has been accordingly claimed to capture better than other approaches how language develops, and particularly, how language breaks down in many clinical conditions (Benítez-Burraco, 2020). However, while some have cited the success of systems biology as evidence to support emergentism (Mazzocchi, 2011; Noble, 2008), such methodological constraints warrant emergentism only on strong interpretations (see Tabatabaei Ghomi, 2023 for an example). On alternative weaker interpretations, they are compatible with the reductionist view that biology as well as language are, in principle, decomposable to their underlying mechanisms, and ultimately to chemistry and physics.

The alternative strong interpretation of indecomposability is a promising basis for an emergentist programme, although linguists have generally not pursued this path. One problem with this alternative is that if one sticks with the reductionistic metaphysics that is prevalent among scientists, it is hard to justify that any phenomenon is indecomposable in principle. If everything

is ultimately a result of physical atoms and their interactions, the view called generative atomism (Humphreys, 2016a), why would a cognition with sufficient resources not be able to decompose everything to those atoms and interactions? Atoms in this context refer to the most fundamental units of physics, whatever they are, and they do not necessarily correspond to chemical atoms as we know them. The idea is that a strong enough cognition would be able to decompose the world to the most fundamental physical units.

Hypothetically, supporters of the emergentist programme could reject generative atomism, and claim *in principle* indecomposability of language. But rejecting generative atomism detaches the emergentist programme from the mainstream scientific worldview that takes everything to be ultimately governed and explained by physics. This is a cost that most language emergentists are not ready to pay. This challenge is not peculiar to language emergentism; it is facing many modern emergentist approaches. For instance, consider Dupre's *promiscuous realism* as a well-known emergentist view. Roughly speaking, Dupre posits that there are many ways one can "cut" the world into legitimate, real, and objectively identifiable natural kinds and therefore, the special sciences such as biology that describe the non-physical "cuts" cannot be reduced to physics that cuts the world in a different way. It follows that science as whole is not unifiable (Dupré, 1993, 1996). The challenge, as Ereshefsky notes, is that "[i]f… the world is as disunified as Dupre maintains, then many current research programs are based on false metaphysics." (Ereshefsky, 1995:143).[17] This is not a knock-down argument against Dupre's promiscuous realism or any other emergentist view. After all, a hardcore emergentist may suggest science needs a paradigm shift in its core metaphysics. What we want to emphasize though is that those who are not ready to embrace such a metaphysical paradigm shift will have a hard time defending their strong emergentist views. This applies particularly well to the majority of language emergentists.

One possible way for language emergentists to claim indecomposability of language to atoms without rejecting the physicalist worldview is to place the locus of indecomposability and hence, emergence in the physics itself. If physics itself warrants some sort of indecomposability and emergence, and if we can somehow connect language to that physical emergence, then claims of language indecomposability and emergence might not be physically unorthodox after all. Language emergentists' best bet here is quantum physics which, some believe, warrants indecomposability and emergence. For instance, Humphreys (1997, 2016c) has claimed that the quantum physical

[17] Ereshefsky is generally sympathetic to Dupre's ideas. For instance, after highlighting the point that if we accept that the world is disunified the underlying metaphysics of many research programmes would be wrong, he continues, "[o]ne area where this may be true, and where it could have dire social consequences, is the current human genome project." (Ereshefsky, 1995:143)

fusions result in emergent entities that are indecomposable to their precursors before fusion. One example are the properties of electrons after they form a covalent bond, which are not decomposable and explainable in terms of their properties before that fusion. Humphreys calls this *transformational* emergence. Emergentist linguists might want to somehow connect the indecomposibility and emergence of language to such quantum transformations.

The views diverge on the possibility and the correct physical and philosophical interpretation of quantum physical emergence (French & Redhead, 1988; Humphreys, 2016d; Joos, 2006; Kronz & Tiehen, 2002; Ladyman & Ross, 2007; Tegmark & Wheeler, 2001). In-depth discussion of that topic is out of the scope of this article. Let us grant, for the sake of argument, that quantum physics warrants some sort of indecomposability. Such a quantum physical indecomposability will form a basis for emergence of not only language, but everything that is built on quantum physics, which according to modern scientific reductionistic view includes everything in the world. This forms a strong basis for those emergentists who wish to establish emergence as a universal phenomenon (for example, see Cordovil et al., 2022). But it does not provide sufficient basis for language emergentists who seemingly claim that language belongs to the exclusive club of emergent phenomena because of its peculiar qualities. They need more than mere universal quantum mechanical indecomposibility to make a particular case for emergence of language.

Perhaps one path to make a special case for emergence of language based on quantum physical indecomposibility and emergence is through the sort of ideas posed by Gallistel and King (2010). Gallistel and King suggest that the brain functions similar to a digital computer. By analysing the information processing constraints that the brain computer would face, they suggest that the computations must take place at the molecular level. Pushing this idea further, one might hypothesise that brain is a quantum computer and language, as one output of this computer, is as instance of quantum computing. In this picture, any sort of emergence that happens at quantum physical level has the potential to directly permeate to language. The problem with this line of thought is that it involves a few big steps, all of them on shaky grounds. First, as noted, the general idea that theories of quantum physics support emergence is debated (Bitbol, 2007; Castellani, 2002). Even the supporters widely disagree on the correct way to understand quantum indecomposibility and emergence (Batterman, 2001; French & Redhead, 1988; Humphreys, 2016d; Joos, 2006; Kronz & Tiehen, 2002; Ladyman & Ross, 2007). Second, the claim that the brain is a digital computer faces important opposition (Penrose & Gardner, 1989). In particular, Gallistel and King's (2010) characterisation of brain as a digital computer is criticised for not being supported by the best available evidence (Donahoe, 2010). Third, even if we accept that brain is a computer in general, we still need additional evidence before we can accept that it is a quantum

one, and in fact, there are some arguments and evidence to the contrary (Litt et al., 2006; Tegmark, 2000). Finally, after all these steps, there still remains the big challenge of showing how quantum emergence in brain quantum computing permeates to language and affects its shape. The supporters of this direction have yet miles to go. They might eventually succeed, but until then, we set the idea of basing language emergence on quantum emergence aside.

Emergentist linguists may give up on indecomposability of language and take a completely different direction. They might accept that language is decomposable in principle, but contend that decomposing language to non-linguistic infrastructures would need an impossibly large amount of epistemic resource that is ever beyond human epistemic capabilities, making language unexplainable by its non-linguistic infrastructures. On this approach, emergence of language would be an instance of epistemic, as opposed to metaphysical, emergence (Chalmers, 2008; Wilson, 2021). As noted in §2, metaphysical emergentists see distinctive characters of the emergent as genuine, real features of emergent phenomena. Epistemic emergentists, on the other hand, regard the distinctive characters of the emergent phenomena not as their genuine features, but as side effects of our limited knowledge and epistemic abilities. As also noted in in §2, epistemic limitations could be deemed surmountable or insurmountable, and it is only the insurmountable limitations that provide a proper basis for a substantial theory of emergence. In the case of language, if the limitations are surmountable, language emergence would be only a passing placeholder until linguists overcome the limitations. But if they are deemed insurmountable, one would be obliged to acknowledge that language cannot be ever explained by its lower-level constituents and thus, will remain forever *epistemically* emergent.

A linguist who wants to pursue this path should put forth arguments to establish that the epistemic barrier against decomposing language to its infrastructure and precursors is, in fact, insurmountable. For that, our linguist can draw, for instance, on computational principles falling under Chomsky's third factor. Supporters of computational theories of emergence have argued that computer-theoretic limitations put a mathematical and hence, insurmountable barrier against full explanation and accurate prediction of emergent phenomena (Bedau, 1997, 2008; Huneman, 2012). In the same line, our linguist could say that third factor computer-theoretic limitations imposed on language guarantee that language would remain ever unexplainable and unpredictable by its infrastructure and precursors. Actually, we already saw above that some advocates of language emergence such as Bates et al. (1998) and Deacon (2003) seem to move in this direction when they cite the mathematical character of language as the reason behind its unpredictability by the non-linguistic infrastructure and precursors.

That said, computational theories of emergence are open to serious objections (Tabatabaei Ghomi, 2022). But let us suppose, for the sake of argument, that some computer-theoretic characters of language establish an insurmountable epistemic barrier that guarantees unexplainability of language in terms of its infrastructure and precursors. Under this supposition, one would find any attempts to explain language by non-linguistic infrastructures to be useless. The emergentist programme that would follow would be totally different from the emergentist programme in its present shape, or the general approach in biolinguistics, here loosely understood as the study of the biological aspects of language. Once one accepts that language is, metaphysically or epistemically, forever unpredictable or unexplainable by its underlying non-linguistic infrastructure and precursors, one has to consequently redirect one's efforts to implement theories of language that are essentially restricted to the level of language itself. But almost all emergentists such as Deacon (2005), MacWhinney (2015), and O'Grady (2008) seem to be going to the opposite direction, as they try to explicate language in terms of its biological bases. On the other hand, if the emergentists see the unpredictability and unexplainability of language as a surmountable limitation, that is, only as a passing side effect of the current limited theories and tools of linguistics, then they are justified to stick to emergentism only until advances in our understanding of the biological (and perhaps physical too) foundations of language renders emergentism obsolete. This view diminishes the emergentist programme to a temporary palliative until reductionistic biolinguistics find the main remedy.

3.4.3 Causal distinctness

One other type of independence attributed to language components is their causal or functional distinctness. In particular, in Deacon's (2003) theory of emergence, language components can do things that the non-linguistic infrastructures cannot. For example, language conveys complex thought in a way that none of its precursors and no part of its infrastructure could. To understand this causal claim more accurately, we need to take one step back from language and look at Deacon's general theory of emergence.

Deacon (2003) claims that all instances of emergent phenomena can be explained as various types of what he calls *topological reinforcement* or *amplification in pattern formation*, where amplification means recurrent superimposition of the same forms. For him, all emergent phenomena are consequences of some sort of a circular causality such as a negative feedback loop. Configurations that result from circular causality get amplified, meaning that the same topology is repeated over larger space and time scales. In some cases, the amplified topologies converge into specific attractor patterns. Deacon thinks that these patterns can have their own causal capacities beyond

the causal capacities of the infrastructures and thus, can perform unique functional roles. He classifies emergent phenomena into three levels of increasing complexity, but at the core of all these three levels are the patterns resulted from amplification by circular causal interactions (Deacon, 2007).

It is hard to construct a consistent and substantial concept of emergent causation from Deacon's various descriptions of these causes. But here are a few key points about these emergent, or, as Deacon (2007) calls them, "configurational" causes. These causes do not originate from new fundamental physical laws (Deacon & Cashman, 2012). Rather, they originate from the particular way that matter and energy are organised (Deacon, 2003). The particular organizations, or topologies of constituents of a system, constrain the way that energy can flow in that system, and give that system new causal capacities (Deacon, 2007).

On a lightweight interpretation of these causal claims Deacon's theory would not be a theory of emergence. Think about your bicycle. The various parts of the bicycle have the causal property of being able to carry you only because they are put in a certain arrangement and design corresponding to a bicycle. If the parts were not in that arrangement, they could not carry you. Therefore, in a way, it is the topology of the bicycle that gives it the causal and functional capacities of being a bicycle. Nonetheless, what actually bears your weight and moves you around is not the topology, but are the parts that form that topology. The causal capacities of a bicycle are not really above and beyond the properties of the parts of the bicycle. In fact, any system that is comprised of more than one element shows the causal capacities that it shows because of the particular way that its elements are organized. But this is not enough to claim that the system has some configurational causal powers beyond the elementary causal powers of its parts. Applying Deacon's terminology to the case of language, if we understand new linguistic causal powers the same way that we understand the causal capacities of being a bicycle, then these putative new causal powers do not make language an emergent phenomenon. A non-emergentist linguist couldn't agree more that, similar to her bicycle, the non-linguistic infrastructures can have linguistic functions only when they are interacting in some particular way. But this acknowledgement does not make her subscribe to any emergentist thesis.

It is only on a stronger understanding that Deacon's theory becomes a substantial theory of emergence. On this stronger understanding, the arrangement of matter and energy in a certain topology results in properties that are completely different from the sum of the properties of matter and energy outside of that topology. Think about life, as an example. One might claim that once the biological parts get arranged in certain organismic topologies, the organism as a whole

shows the property of life and its associated causal capacities such as reproduction that are distinct from the properties and causal capacities of the biological parts. Deacon's theory becomes a substantial theory of emergence only with a such strong understanding of emergent causal properties. Such a strong understanding is also more consistent with Deacon's (2003) claim that his account is an instance of ontological emergence, which can be regarded as just another term for metaphysical emergence, while in some other places Deacon seems to distance himself from such strong interpretations by emphasising that emergent causal powers are nothing but manifestations of constraints on fundamental causes (Deacon & Cashman, 2012). A non-emergentist linguist might be fine with metaphoric references to linguistic causes as theoretical placeholders. But only a truly emergentist linguist can accept those causes as properties peculiarly belonging to components of language. Therefore, with the strong interpretation, Deacon's ontological emergence can form the basis for a truly emergentist programme for language that is substantially different from non-emergentist programmes. The resulting programme will also be different from the programmes promoted by negative accounts of language emergence such as O'Grady's emergence or the MP. Unlike those accounts which emphasise the dependence of language on non-linguistic factors, the strong interpretation of Deacon's emergence promotes a programme that emphasises the independent causal and functional roles of language. Also, in contrast to the emergentist programme promoted by MacWhinney in which the concept of emergence plays no significant role, emergence sits at the heart of this alternative hypothetical programme by bestowing on language all its distinctive causal and functional capacities.

Deacon's account, however, faces important challenges, such as Kim's exclusion argument, the problem of realisers for emergent causes, and (which is very relevant for our concerns here) limitations in its applicability to language. We discuss the applicability limitations and the problem of realisers in this section and leave the discussion of Kim's exclusion argument to section 5.

Deacon promotes his theory as an all-encompassing theory of emergence. But because he takes pattern amplification to be the core source of emergence, his theory is applicable only to those cases where emergence happens because of pattern amplification. Nonetheless, many classic examples of emergence do not involve the sort of pattern amplification that Deacon associates with emergence and, thus, fall out of the scope of Deacon's theory. Game of Life (Berlekamp et al., 1982) is one such example. Game of Life is a special computer simulation setup in which a random configuration of dead (coloured white), and alive (coloured black) cells following a few simple rules, eventually evolves into elaborate structures of dead and alive cells with unexpected regularities. These structures are so conspicuous and stable that are identified and characterised as different *species* by the community of Game of Life researchers. Game of Life species are famously

used as models of epistemic (Bedau, 1997) as well as metaphysical emergence (Dennett, 2008; Wilson, 2021). Formation of Game of Life species, however, does not involve pattern amplification as described by Deacon. The role of pattern amplification is also not clear in formation of prime real-world examples of emergence such as mental phenomena. For instance, what is the role of pattern amplification in formation of the introspective feeling of pain? And most importantly for our present purpose, the role of pattern amplification in the emergence of language is not clear. For instance, what role does pattern amplification play in recruiting non-linguistic machineries for linguistic articulation? Therefore, it is not obvious if Deacon's theory can explain emergence of language - or that language can be truly regarded as an emergent phenomenon, should we adopt Deacon's theory as our theory of emergence.

Deacon himself does not fully clarify how his general theory of emergence applies to the special case of language. Language appears only passingly in the conclusion section of one of Deacon's (2003) older theoretical papers, and his later works that are focused on emergence of language (Deacon, 2005, 2014) are also missing a clear connection between his general theory of emergence and his views on emergence of language. These works are primarily concerned with describing various sublinguistic levels and the connections between them, but they do not clarify how the core of Deacon's theory of emergence, namely, recurrent superimposition of the same forms, generates emergent linguistic causes from these sublinguistic levels.

All that being said, let us grant, for the sake of argument, that Deacon's theory can somehow be applied to the case of language emergence. Still, an important problem is that a key feature of Deacon's theory of emergence is that he denies that emergence is about new "things" coming to existence and claims, instead, that what emerges are mere properties, such as higher-level causal capacities and functions which result from low-level causes constrained in specific patterns: "[a]mazing new properties have been, and are being, emerged, and there is nothing new being added. There is no thing new" (Deacon & Cashman, 2012: 204). If this is true, we should then acknowledge that there is no such thing as language, only functions that are performed by language. One important function of language, or causal capacity if you like, is conveying thoughts. So, following Deacon, we should say that there exists the function of conveying thoughts as a real fact of nature, but there is no such thing called language. Eventually, we would end up with a linguistics that does not acknowledge the existence of language to begin with. Philosophically speaking, the problem here is that the emergent functions and properties cannot exist in a vacuum. They need realisers; things that show those functions and properties.

Deacon proposes a solution to this problem. He associates the emergent functions and properties with topologies, that is, the way the lower-level components are arranged. So, for him, it is the special arrangement of the infra-linguistic components, i.e., their topologies, that realise the emergent linguistic functions. But taking the topologies to be the realisers of emergent causal powers has an important and potentially uncomfortable consequence that should not be ignored. Topologies are non-physical, abstract concepts. One who ascribes causal powers to topologies implicitly claims that abstract concepts can have causal powers. This is a hefty metaphysical claim with a lot of philosophical and scientific consequences that should not be overlooked. For instance, this commitment is opposed to the widely accepted view among physicalist scientists that the only causal powers in the world are those of physical matter. Ultimately, Deacon's account would no longer be compatible with the prevalent scientific materialist physicalism about causes.

A dedicated emergentist linguist can bite the bullet and reject that the only causal powers in the world are those of physical matter. Such a linguist can subscribe to the alternative metaphysics of ontic structural realism (Cordovil et al., 2022; Ladyman, 1998; McKenzie, 2017; Santos, 2015), according to which structures are real natural objects capable of exerting causal effects. One who accepts this metaphysics can safely attribute causal powers to structural topologies. But accepting ontic structural realism instead of the standard materialist physicalist metaphysics is a significant deviance from the widely held metaphysics among scientists, and language emergentists such as Deacon do not seem to be comfortable with it.[18] Language emergentists seem to be half-hearted when it comes to making difficult decisions that are required for constructing a philosophically coherent view of language emergence.

What if we construct a modified version of Deacon's theory by accepting, contra Deacon, emergence of language and its components as emergent "things" and ascribing emergent causal powers and properties to those entities? Such an account would be exposed to Kim's exclusion argument (Kim, 1992, 1998, 1999), which we discuss in detail below. Before going to Kim's argument, however, we will discuss the last way that language can be described to be distinct and independent with respect to its underlying non-linguistic infrastructures, i.e. their organismic life. This type of property is also exposed to Kim's exclusion argument.

[18] Deacon emphasizes time and gain that he does not want to add anything to the physical causes as we know them. For instance, he writes: "No novel types of physical causes are evoked by this conception of emergence, only novel types of configurations and what might be described as 'configurational causes.'" (Deacon, 2007:109)

3.4.4 Linguistic organismic life

As noted, a final way that emergentists have described the distinctness and independence of language with respect to non-linguistic infrastructures is by ascribing an organismic life to the components of language (Frank, 2008). One of the pioneers of this view was August Schleicher, who on top of being a linguist, was also a botanist and a gardener, and ended up merging views from all his different fields of interests by describing language as a living organism (Austin, 2021). This view has been pursued by modern theorists who characterize language components as species that behave somehow like genes or viruses (Ritt, 2004). One of the most elaborate examples is Piatelli-Palmarini and Uriagereka's (2004) description of the evolution of language morphology by reference to the corresponding details in the evolution of viruses. More broadly, language items can be seen as instances of so-called *memes*, the cultural units of Dawkins' memetic theory, which he claims spread in societies via the process of imitation, similar to how genes spread in biological populations (Dawkins, 2016). According to this organismic view, the components of language follow their own life and evolutionary dynamics. Importantly, this evolutionary dynamics of language items should not be confused with the evolutionary dynamics that happens on the biological level. The evolutionary dynamics of the diverse language species (and of language as such) happens in a linguistic ecosystem shaped by social and cognitive factors (Frank, 2008), and the fitness value in this ecosystem is determined by how well the species serves its functions: cognitive, communicative, interactional, etc. This picture of language items (and of language) evolving in response to its environment finds support in the growing evidence that points to the effect of social transmission in shaping language features (Kirby et al., 2015), and more generally, on the effect of the social (but also the physical) environment on language properties (Lupyan & Dale, 2016).

The question here is whether ascribing life and evolutionary dynamics to linguistic components is a metaphoric analogy, or a realistic claim. Sometimes one single text reflects both understandings. For instance, Piattelli-Palmarini and Uriagereka say that "[w]e have *likened* morphology to a virus" (Piattelli-Palmarini & Uriagereka, 2004: 357, the emphasis is ours), seemingly going towards the metaphoric understanding, but they also write, in the same article, "natural languages are rich 'objects' to which a variety of characterizations truthfully … apply" (Piattelli-Palmarini & Uriagereka, 2004:342), swinging to the realistic understanding. These conflicting inclinations sometimes show themselves even in one single sentence: "we think our metaphor is productive and worth pursuing to its several interesting consequences. Indeed, there are reasons to believe that it may be more than just a metaphor" (Piattelli-Palmarini & Uriagereka, 2004:365). The issue of realistic vs metaphoric understandings of the organismic characters of

language is a matter of debate (Austin, 2021), and we do not want to take sides here. So, we consider both options.

On the metaphoric interpretation, speaking of life and the evolutionary dynamics of linguistic components is merely an expressional tool. It does not show any genuine property on the language level and, therefore, it does not make language an emergent phenomenon. On the other hand, with a realistic interpretation, the linguistic life and evolutionary dynamics of linguistic components are genuine properties that are completely absent from the non-linguistic infrastructures and warrant taking language as an emergent phenomenon. Claiming that language components are real species means that they possess real organismic functions and causal capacities within the linguistic ecosystem which their non-linguistic infrastructure and precursors completely lack. These distinctive functions and causal capacities expose the emergence of language to Kim's exclusion argument discussed below.

3.5 KIM'S EXCLUSION ARGUMENT

Kim's exclusion argument (Kim, 1992, 1998, 1999) is one of the most important challenges facing a wide range of different accounts of emergence. The first premise of the argument states that genuine scientific kinds are identified by their causal powers. If some kind of thing is not associated with any causal capacities, then it is not a genuine scientific kind. In the case of language, if a language component does not play any causal role, then it is not a genuine scientific kind and should be eliminated from scientific theorising. This premise of the argument might not be a big problem for some current emergentist approaches to language. As we already saw above, there are some emergentists such as Deacon who (at least on some interpretations) associate emergence with higher level causal powers. These theories of language emergence pass through the first step of the exclusion argument. In fact, even theories that associate "dysfunctions" to language (Piattelli-Palmarini, 1990) can pass through, as long as they associate language with whatever sort of function as a capacity to causally affect the state of the things, no matter how.

The next premise of the exclusion argument is that the lowest physical level is causally closed. This means that any lower-level physical phenomenon has an exclusively lower-level physical cause. Another premise of the argument is that any higher-level emergent phenomenon ultimately supervenes on the lowest physical level. *Supervenience* is a technical term, and it means that higher-level phenomena are necessarily realised by lower-level physical infrastructures, and there cannot be any changes in any higher-level phenomenon unless it is accompanied by a change in the phenomenon's underlying lower-level realisers. Now suppose that a higher-level phenomenon

H_1 that is realised by a lower-level physical structure L_1, causes a higher-level phenomenon H_2 that is realised by a lower-level physical structure L_2. Because of the causal closure of the physical, L_1 causes L_2 with no input from H_1. But since H_2 is realised by L_2, L_1 would be causally sufficient to produce H_2, and there is no need for any higher-level causation between H_1 and H_2. Introducing the higher-level causation results in causal over-determination and is therefore unacceptable. It follows that the higher-level phenomenon H_1 cannot have causal powers and therefore, is not a genuine scientific kind.

The conclusion of Kim's exclusion argument is that postulating emergent entities such as language components is either unnecessary, or problematic. If these components do not have any causal power, then they are not genuine scientific kinds and they are unnecessary for theorising. But if we claim that they do have causal powers, then we face the problem of causal over-determination. Going back to our main discussion, we saw above that some emergentist claims such as Deacon's emergent causes and organismic functions of linguistic elements can have two lightweight, and strong interpretations. If Kim's exclusion argument is sound, then the lightweight interpretations do not add anything substantial to linguistics, and the stronger interpretations are in conflict with the prevalent view that everything in the physical world is governed exclusively by physical causes.

Kim's exclusion argument is a general anti-emergentist argument that is facing any positive account of emergence, in linguistics or other contexts. As such, there is a vast philosophical literature on this argument from which the emergentist linguists can potentially borrow counterarguments and formulate their responses accordingly (for some examples see Kallestrup, 2006, and O'Connor & Wong, 2005). That being said, let us see how Deacon as an example language emergentists have reacted to Kim's argument.

In reaction to Kim's exclusion argument, Deacon and Cashman write:

> Kim's criticism is almost certainly right, given the assumptions of this part/whole conception of causality, but rather than undermining the concept of emergence, it suggests that certain assumptions must be wrong. A number of contemporary emergence theorists have answered this criticism by arguing either that the notion of compositionality is not as simple as Kim and most eliminativists have assumed, or that the assignment of causal power to the lowest possible level of compositionality is a mistake. Determining simple decomposition into proper parts is, for example, a problem when it comes to the analysis of organism. … With cessation of the life of the organism – that is, catastrophic dissolution of these critical reciprocal

> interrelationships – the components rapidly degrade as well. So their properties are in part derived from this synergistically organized higher-order dynamic. (Deacon & Cashman, 2012:197)

Despite his general attitude of clinging to physicalism, it seems that in reaction to Kim's argument, Deacon tends to adopt a stronger form of emergentism and accept that the emergent properties such as emergent causal powers are at least partly sourced by "synergistically organized higher-order dynamic." Language, therefore, could be taken to possess functions that are genuinely distinct from the causal powers of the non-linguistic infrastructures and, ultimately, the causal powers of the physical particles. This is a solid response to Kim's argument. However, the linguists who opt for such a response should be ready to accept some form of non-physicalism, which will make the resulting emergentist programme deviate from the common physicalist metaphysics of science. They should accept emergent causes that cannot be fully analysed to, and thus dissolved by the physical causes. This is equivalent to accepting that the natural world is at least partly governed by nonphysical causes which is against the prevalent scientific physicalism. An ardent emergentist might be ready to accept all these consequences. In fact, this seems to be the correct choice for someone who is wholeheartedly committed to emergentism. But our review of the emergentist programme's literature suggests that many if not most emergentists have not adhered to, and would not be in fact ready to accept these conclusions. Despite their emergentist banner, the majority of them seem to stay fully committed to the prevalent reductionist and physicalist metaphysics of science. As O'Grady characterises their thesis, emergentists think that language is best explained by reference to more basic non-linguistic infrastructures. They are not ready to adopt a fully emergentist position that regards language as an irreducible, metaphysically distinct phenomenon. In fact, on some occasions they explicitly denounce such a full force emergentist approach. Even Deacon (2003:274), for instance, writes, "[t]he concept of emergence probably has gained its worst reputation when it has been used in a primarily negative sense—to point to something missing in reductionistic explanations". Philosophers such as Sober (1999) have been wary of this kind of half-hearted emergentism and have pointed that physicalist anti-reductionists who hold that the only inherent causal powers are those of the physical lower-level are anti-emergentists at heart, no matter what they say. We showed that this kind of half-hearted emergentism does not add much to linguistics. And on the other hand, by clarifying the corollaries of wholehearted emergentism, we showed that it is probably not an attractive alternative for many language emergentists.

3.6 CONCLUSIONS

In this paper, we examined the concept of language emergence in linguistics and neuroscience as a special case of scientific deployment of the concept of emergence. We discussed various philosophical challenges to different understandings of the concept of language emergence along with some possible solutions to those challenges. Although we focused on the emergentist programme within linguistics, one can generalise many lessons drawn to emergentist programmes in other disciplines.

Through our analysis, we repeatedly came to junctions where one could adopt a strong or a weak understanding of the concept of emergence. We saw that on weak understandings, the emergentist programme would not be that different from non-emergentist programmes. It is on the strong understandings that the emergentist programme finds its unique character, but that unique character comes at a hefty cost. Adopting strong stances on language emergence comes with consequences such as accepting causality of abstract concepts, endorsing non-physicalism, or rejecting physical causal closure. Emergentists may well accept these consequences and in fact, this will contribute to developing a philosophically coherent theory of emergence. However, a programme based on such views would deviate from the physicalist reductionistic metaphysics held by the majority of the scientific community. Obviously, scientific facts and truth in general are not determined by majority vote and so, the strong emergentist choices cannot be ruled out simply because they are not popular. Nonetheless, our review shows that many who hold the banner of language emergentism would probably avoid choosing those options.

A hypothetical ardent language emergentist who chooses to pick the strong options has two choices. One choice is to let go of the common physicalism and opt for a programme based on metaphysical emergence. The other choice is to opt for a programme based on epistemic emergence and regard language as epistemically ever-unexplainable by its non-linguistic infrastructure. We argued that both the metaphysical and the epistemic approaches would result in a kind of hypothetical emergentist programme that is different from the current so-called emergentist research and the general trends in linguistics, and biolinguistics and neuroscience of language more specifically. Such a hypothetical programme would be in opposite direction to what O'Grady suggests as the unifying theme of all current so-called emergentist approaches which is the claim that the phenomena of language are best explained by reference to non-linguistic factors. Such a hypothetical emergentist programme would either take the linguistic components to have independent irreducible metaphysics of their own, or take explaining them in terms of their underlying non-linguistic bases to be forever beyond our epistemic capacities. A programme with

such theoretical commitments would focus on explaining language in terms of linguistic laws and regularities, acknowledging that those laws and regularities themselves will remain forever unexplained by reference to their underlying non-linguistic factors. The resulting programme would involve much less reductionistic mechanistic explanations, and much more holistic descriptions.

Language emergentism in its current form, as much as it deviates from a purely technical claim towards a philosophical position, wants to keep the reductionistic physicalist metaphysics untouched, and at the same time embrace emergentism. Our close analysis of the programme, however, shows that such a hybrid is not viable. We showed that there are junctions where one has to choose between emergentism or reductionistic physicalism, and those who want to go both ways end up nowhere. The core lesson of our analyses that guides this conclusion can be summarised in the following adage: Half-hearted emergentism is expendable. Wholehearted emergentism is expensive.

Chapter 4
The minimum set of necessary ontological commitments for structural metaphysical emergence

ABSTRACT

The literature on metaphysical emergence, the idea that systems as a whole can have standalone metaphysical reality and higher-level causal powers, is usually about the ontological, scientific, methodological, and other arguments for or against metaphysical emergence. In this chapter, I turn this common practice on its head by taking structural metaphysical emergence for granted and searching for the ontological picture of the world that fits metaphysical emergence. I begin by clarifying the specific version of metaphysical emergence that I assume in terms of two claims, one metaphysical and one causal. I then survey a number of representative structuralist accounts of emergence to find the minimum set of ontological commitments necessary for maintaining the metaphysical and causal claims of metaphysical emergence. I conclude with a list that has three elements: (A) structural realism, (B) structural causation, and (C) the condition of downward percolation.

4.1 INTRODUCTION

There are many instances where systems show characters and functions that are apparently distinct and autonomous with respect to their constituent parts. Examples range from mysteries such as intentional mental causation to miseries such as flow dynamics of traffic jams. Metaphysical emergence is the idea that, at least in some of such cases, the distinct characters and functions of the system as a whole belong to the whole itself as a metaphysically distinct entity and cannot be explained away in terms of the properties and functions of the parts (Wilson, 2021). Many have argued in support of metaphysical emergence (Chalmers, 1996; Deacon, 2007; Humphreys, 1997, 2016d; Huneman, 2008; O'Connor & Wong, 2005; Santos, 2015; Tabatabaei Ghomi, 2023; Walsh, 2015a; Wilson, 2021, 2015; Wimsatt, 2000). Here I am not concerned with adding yet another argument. Instead, I take metaphysical emergence as my starting point and ask how to adjust one's grand view of the world accordingly.

In particular, I search for an ontological picture of the world that can accommodate metaphysical emergence understood as emergent structural entities with genuine causal effects. By an ontological picture I mean the description of the type of things that exist in our world and the causal and metaphysical relations between them. My search for an ontology for metaphysical emergence goes in the reverse general direction of the emergentist literature that starts from ontological or scientific discussions to conclude that metaphysical emergence is possible. I take metaphysical emergence for granted and ask what ontology befits it. So, my discussions primarily face those who have already accepted metaphysical emergence or at least take it seriously. But the conclusions can be of interest even to those who are inimical towards metaphysical emergence. They might find the resulting ontological picture so bizarrely remote from their firmly held ontological views that they can take it as yet another reason to reject metaphysical emergence.

The most common ontological view that is tacitly assumed among many scientifically minded thinkers is what Humphreys (2016a) calls generative atomism. In this ontological picture, our world is made of some fundamental particles. Let us call them, whatever they are, *atoms*. Anything else in the world is simply a particular arrangement of these atoms, and therefore, the atoms have metaphysical and causal primacy over everything else. Everything is, at least in principle, reducible to atoms. As Humphreys points out, this atomist reductionistic ontology is incompatible with metaphysical emergence which attributes distinct metaphysical existence and causal roles to non-atomistic emergent entities. One cannot commit to this ontology and, at the same time, accept metaphysical emergence. Yet, some emergentists have tried to reconcile generative atomism and metaphysical emergence, with the main incentive of keeping emergentism compatible with the mainstream scientific worldview. I will critically discuss some of these accounts and argue that the ontology of generative atomism is irreconcilable with metaphysical emergence.

But what are the alternative ontologies that a metaphysical emergentist can adopt? In particular, what absolutely necessary ontological claims must a metaphysical emergentist commit to? Answering this question is the primary pursuit of this chapter. By looking deep into the claims of metaphysical emergence, as well as a number of representative structuralist accounts of metaphysical emergence, I construct a list of the minimum ontological commitments necessary for structuralist theories of metaphysical emergence. I conclude with a list that has three elements: (A) structural realism, (B) structural causation, and (C) the condition of downward percolation.

I begin in section 4.2 by covering some background, clarifying my assumptions, and specifying what I mean by a structuralist view of metaphysical emergence. The necessary ontological commitments begin to show their outlines from this very first step. In sections 4.3 and 4.4 I

critically analyse a few representative structuralist theories of emergence to find the ontological elements needed for such theories to be substantial, coherent, and robust. In section 4.5, I compile my minimum set of necessary ontological commitments for structural metaphysical emergence. I also give an example of an ontology that has all the elements of this set and, therefore, is suitable for structural metaphysical emergence. And section 4.6 will be my conclusion.

4.2 CLAIMS OF METAPHYSICAL EMERGENCE

As noted, I take metaphysical emergence as my premise. But there are many understandings of metaphysical emergence, and so it is essential to clarify what I mean by the term. We can summarise metaphysical emergence in terms of two claims, one metaphysical and one causal. In this section, I clarify these claims, define metaphysical emergence in terms of metaphysical and causal indecomposability, and specify the particular type of structuralist metaphysical emergentism that I take as my starting point. I will explain the relationship between the two claims of metaphysical emergence and the view of emergent phenomena as structural entities. In particular, the discussions of this section clarify the ontological concepts of real structures and emergent causal capacities, and set the ground for the structuralist ontology that I propose and support in the subsequent sections.

4.2.1 The metaphysical claim

> **The metaphysical claim**: There are systems that, as a whole, form real entities that are metaphysically distinct from their constituent parts. [19]

An immediate question follows the metaphysical claim aiming at the nature of the emergent entities. If emergent entities are not metaphysically equivalent to the sum of their parts, then what are they? What is the stuff that makes the emergent entities?

Various forms of dualism, such as vitalism or mind-body dualism (Descartes, 1998; Driesch, 1905) and non-structuralist accounts of metaphysical emergence (Chalmers, 1996; O'Connor & Wong, 2005) suggest that the emergent entities are made of their own peculiar and potentially nonphysical substance. In contrast, structuralist metaphysical emergentists claim that the emergent entities are *structures* (more on this below) (Auyang, 1998; Deacon, 2006; Morgan, 1925; Santos,

[19] I frame the claims in part-whole terms. But it is worth noting that there are some accounts of metaphysical emergence that are not expressed in part-whole terms. Most famously, Humphreys (1997, 2016) has suggested that even non-composite objects can show emergence. I'd rather avoid Humphreys' view and stick to the part-whole notion of emergence because, first, Humphreys' account has been subject to important objections from other metaphysical emergentists (Wong, 2006), and second, the part-whole view is the dominant view among various theories of metaphysical emergence (Wilson, 2021).

2015; Wimsatt, 2006). Structuralist views of metaphysical emergence seem to be more in line with science, so they have been better embraced by scientists and scientifically minded philosophers (Anderson, 1972; M. Mitchell, 2009; Rothschild, 2006; Van Regenmortel, 2004). In this chapter, I am exclusively concerned with these structuralist views.

What does it mean to say that emergent phenomena are structures? More broadly, what is a structure? Theoretically, a structure is a non-empty set of relata associated with a non-empty set of relations (Frigg & Votsis, 2011).[20] Non-relational properties of relata can also be included in the set of relations as relations that do not have a second relatum. A structure so defined can be seen as an abstract concept belonging to mathematics and, as such, be as real as any other mathematical object can be (S. Shapiro, 2000). Metaphysical emergentists who describe emergent entities as structures, however, go beyond that and claim that there are emergent structures that are real occupants of nature. They are *structural realists*. By *realism*, I refer to the concept as understood in the context of the realism vs anti-realism debate in the philosophy of science (Godfrey-Smith, 2003). Real entities exist in nature as objective realities that are, in principle, objectively identifiable, as opposed to non-real entities that are merely functional or instrumental.

The concept of real structures should not be surprising for a scientifically-minded philosopher. Scientists routinely identify, classify, and model structures in the context of various sciences. In physics, mathematical structures are expressed as formulas of different physical properties and interactions. In chemistry, we see molecular structures with their own unique properties and reactions. In biology, we see functional structures at different scales from organelles to ecologies, and so forth. In fact, it is tempting to suggest that the whole business of science is about objectively identifying, classifying, and modelling structures. From the scientific perspective, the idea of objectively identifiable real structures that perform different causal functions is a familiar concept.

Among philosophers, Dennett (2008) has famously argued for the reality of structures, or in his words, the reality of "patterns." For him, structures are real if they can be associated with specific functions and can be objectively identified. Dennett's key idea is that information theory allows us to objectively distinguish structures from random configurations (noise) or other structures. For example, compare a set of 1000 random numbers, with a set of 1000 consecutive numbers starting from 1, with another set containing the first 1000 numbers of the Fibonacci sequence. Suppose you want to transfer these numbers to someone over the phone. For the first

[20] To be more precise, Dewar (2019) has shown that a set theoretic definition of structures that is going to be used by ontic structural realists (to be discussed later in the chapter) should also include the relations among the relations, and so make an "algebra" of relations.

set, you need to read every single one of the 1000 numbers. This set has no structure and therefore, you cannot compress its information content. But the second and the third sets have structures that allow you to compress and transfer their information content with a much shorter message. However, the amount and the type of information you need to transfer the contents of each of the second and third sets differ. These differences give you an objective way to distinguish these three sets.

Ross (2000) has expanded Dennett's idea and has described the metaphysics of our world as a crowded *rainforest* inhabited by countless real patterns of different types and various scales. He calls this *rainforest realism*. According to Ross, a pattern, or in our terms, a structure, is real as long as it satisfies two conditions. The first condition is that the structure should be projectible in at least one physically possible perspective. It is hard to understand what Ross means here. My understanding of what he calls a "perspective" is a dimension along which we can describe some property. A set of dimensions forms a coordinate system. The exact property can be projected in different coordinate systems. For instance, the spatial property of an object can be projected on the Cartesian coordinate system (x, y, and z) or on the spherical coordinate system (θ, φ, and r). Each of these coordinate systems provides different dimensions, or in Ross' terms, perspectives on the spatial property. Ross requires that these dimensions to be physically possible and defines "the set of physically possible perspectives as the set of perspectives available in the possible worlds that are nearby according to physics" (Ross, 2000: 161). Therefore, according to Ross, the coordinate system cannot describe a world fundamentally different from our physical world. For instance, if the physical world is spatially only three-dimensional, a four-dimensional coordinate system is unacceptable. Ross demands that the relations of real structures should be projectible in at least one such physically possible coordinate system.

I think any structure described in the special sciences satisfies this first condition and, thus, many supposedly emergent phenomena. Nonetheless, I suggest relaxing this condition to make it more comprehensive and allow more cases of metaphysical emergence to be included. Ross is explicit that he has set this condition to exclude famous cases in support of nonphysical qualia from his set of real structures. Examples are a colour-blind omniscient neuroscientist (Jackson, 1982) and zombies whose mental experience differs from ours despite having the same neural system (Chalmers, 1996). But we do not need to commit to this exclusive motivation. We can expand the acceptable coordinate systems to all that are logically possible. And if my reader thinks that logical possibility is not well-defined or irrelevant to what is real, we can ask the coordinate

systems to be *naturally* possible, dropping the implicit commitment to physicalism in Ross' first condition.

Ross' second condition is that a real structure should encode information about at least one non-dispensable perspective on a set of events or entities. If we understand Ross' "perspectives" as dimensions, the second condition means that a real structure should encode information about at least one dimension that matters in describing an entity or event. But Ross requires more; the structure should encode this non-dispensable information in an information-theoretically more efficient way than the complete description of those events and entities. We saw an example of such information-theoretic efficiency above in the case of the three sets of 1000 numbers. The non-dispensability condition ensures that Occam's razor is satisfied, so only structures containing relevant information count as real. And the information-theoretic efficiency requirement ensures the objective identifiability of real structures.

Let us look at the familiar example of Game of Life species (Bedau, 2002; Berlekamp et al., 1982; Berto & Tagliabue, 2017; Dennett, 2008) as instances of emergent phenomena that satisfy Ross' conditions for real structures. As a quick reminder, Game of Life is a computational model widely used as a model of emergence. It comprises a two-dimensional lattice of square cells where each cell can be either "alive" and coloured black, or "dead" and white. The cells become or stay dead or alive, following some rules based on the colour of their neighbouring cells. Curiously, some completely random initial configurations of dead and alive cells generate patterns of cells with shapes and behaviours so recognisably ordered and stable that they are identified as special "species" by the community of Game of Life researchers. These patterns are taken to be emergent with respect to their constituent cells. One of the most famous examples of these patterns is the "glider", a species of Game of Life that oscillates in three-timestep periods ($3t$), going through some transformations in its form, and moving one cell diagonally (Figure 4.1).

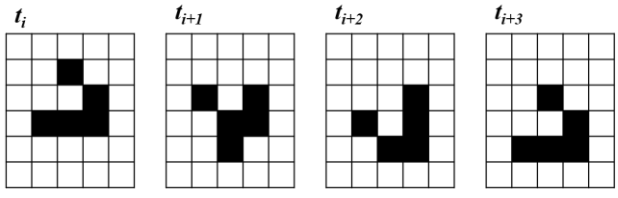


Figure 4.1 Glider in GoL. Glider oscillates and move one cell diagonally with period $3t$.

A glider is a real structure by Ross' definition. It is projectible in spatial and temporal dimensions of Game of Life, and it encodes non-dispensable information about the life and death of its comprising cells efficiently. As Wilson (2021) has pointed out, the number of degrees of

freedom needed to specify the location of live cells in a glider on the level of cells is five (the number of cells comprising a glider) times two (corresponding to the *x* and *y* coordinates), or ten. However, once we recognise the structure of a glider, we need only three degrees of freedom to store the same information; two to specify the *x* and *y* coordinates of one of the cells, say the leftmost, and one to specify the stage of the glider in its periodic transformation. A glider, therefore, satisfies all of Ross' conditions and counts as a real structure.

Ross's definition of real structures and the more relaxed version of it presented above is one way to characterise real structural entities. Of course, a supporter of structural metaphysical emergence does not have to accept this particular definition of real structures. But this way or another, anyone who claims that emergent entities are metaphysically real and simultaneously holds that emergent entities are of structural nature must accept structures into her ontology as real occupants of the world. In other words, she has to embrace some form of *structural realism*. This is the first necessary ontological commitment for any structural account of metaphysical emergence.

4.2.2 The causal claim

The causal claim: The emergent entities have causal powers distinct from their constituent parts' causal powers.

Not all metaphysical emergentists subscribe to the causal claim. Particularly, epiphenomenalists such as Chalmers (1996) do not attribute any causal powers to emergent phenomena, even though they acknowledge the metaphysical reality of the emergent. For instance, they hold that pain is metaphysically real and distinct from the nociception system, but is completely causally inert. Therefore, they suggest that there could be zombies who have a replica of our nociception system but do not feel pain, and those zombies would behave exactly as us, the pain-suffering humans.

Epiphenomenalism is unattractive for two reasons. First, one important observation behind metaphysical emergence is that we introspectively attribute causal powers to our feelings and intentions, and special sciences routinely attribute causal effects to emergent entities. Second, completely causally inert entities cannot be interacted with and, consequently, cannot be studied by science (Kim, 1992, 1998, 1999). Therefore, even if epiphenomenal emergent entities are metaphysically real, they are scientifically uninteresting and dispensable. They could potentially have instrumental value and simplify scientific theorising and modelling. But they can have no "real" place in scientific explanation.

Some metaphysical emergentists accept that emergent entities are causally distinct from their constituent parts. Yet, they contend that this distinctness does not include causes over and above the causal powers of the parts. Instead, they claim that the emergent entities are causally distinct from their parts because they show only a proper subset of the causal effects of the parts and, therefore, have a different causal profile compared to their parts (Shoemaker, 2000; Wilson, 1999, 2010, 2015, 2021). The supporters of this view suggest that having a more limited causal profile warrants taking the emergent entities as standalone metaphysical entities distinct from their parts. The reason is that, by Leibniz's law (Forrest, 2020), identicals must be indiscernible. By modus tollens, if two entities are discernible, they cannot be identical. Therefore, as emergent entities are systematically discernible from the parts by their more limited causal profiles, they must be metaphysically different from their parts.

Consider Entamoeba *histolytica*. According to the proper subset theories of emergence, E. *histolytica* is a metaphysically distinct entity with respect to its constituent parts because we can systematically identify E. *histolytica* by only a subset of the causal effects that its bio-physico-chemical parts exert. For instance, the radioactivity of atoms in E. *histolytica* is irrelevant to E. *histolytica* as an organism. Only a subset of all of the causal effects of those atoms ends up in the causal profile of E. *histolytica* as an organism. Therefore, the thought goes that E. *histolytica* can be systematically distinguished from its parts as a metaphysically separate entity by reference to its more limited causal profile.

Although these proper subset accounts of metaphysical emergence attribute causal powers to emergent entities, I believe that, on a closer look, they are a special kind of epiphenomenalism disguised as otherwise. The causes they attribute to the emergent phenomena are already accounted for when listing all the causes of the parts. So why do we need to count them again and attribute them to the emergent entities one more time? Moreover, all the "power" in the causal powers shared between the parts and the emergent entities comes from the parts, and therefore, the emergent entities seem to be causally dispensable.

Yet more importantly, the proper subset theories do not account for those causal powers of the emergent entities that are ostensibly not present in the causal profile of the parts. Atoms do not reproduce, nor do they escape, nor do they survive. The organisms such as E. *histolytica* reproduce, escape, and survive. Neurons do not intend, feel pain, or wish they were exploring Parque Nacional da Tijuca instead of writing papers. It is the emergent mind that does all these things. Or at least this is how many with emergentist inclinations see the matter.

Putting epiphenomenalism and the proper subset accounts to the side, we get to those theories of metaphysical emergence that attribute unique causal powers to emergent entities (Deacon, 2006; Humphreys, 1997; O'Connor & Wong, 2005; Santos, 2015; Walsh, 2015b). These theories align with emergent entities' apparently indispensable causal roles in special sciences and the introspectively felt causal powers of mental phenomena. Merging these two types of emergent causation results in perhaps the most interesting case of emergent causality, namely, agency.[21]

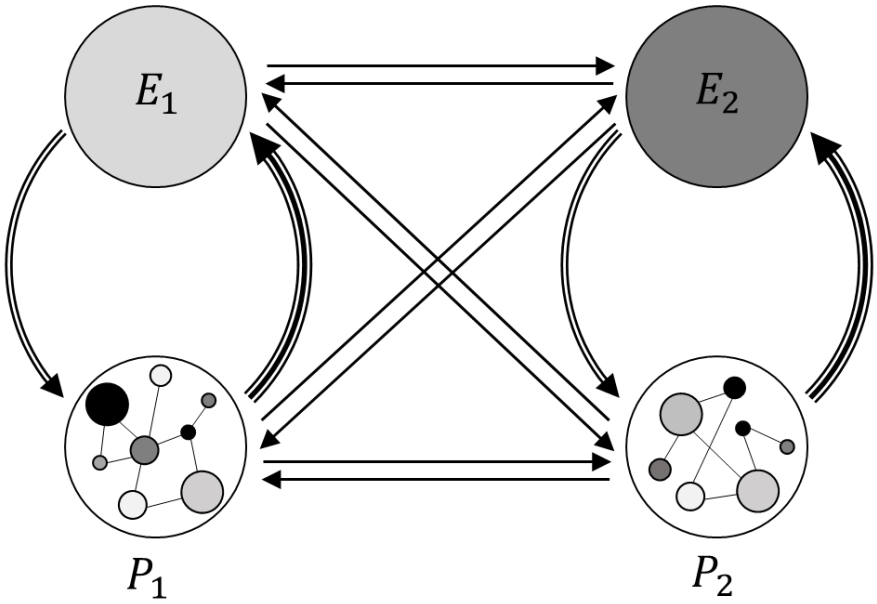


Figure 4.2 Potential causal relations between the parts and the emergent

I graphically list all of the potential causal relations between parts and emergent entities in Figure 4.2. P_1 and P_2 denote two sets of lower lever parts and E_1 and E_2 are their corresponding emergent systems. Each arrow denotes a potential causal relation. The triple-line arrows indicate the causal relation between the parts and the emergent whole, the double-line arrows indicate the causal relation of an emergent whole on its own parts, and the single-line arrows indicate any other causal relations. We can identify six different types of causal relationships in this diagram, $P \Rrightarrow E$, $E \Rightarrow P$, $P \rightarrow P$, $E \rightarrow E$, $P \rightarrow E$, and $E \rightarrow P$. The causal claim can be satisfied by any of $E \Rightarrow P$, $E \rightarrow E$, and $E \rightarrow P$ types of causes, and different theorists of metaphysical emergence emphasise different one of these types. For instance, O'Connor and Wong (2005) emphasise $E \rightarrow P$, while Walsh (2015a) emphasises $E \Rightarrow P$.

[21] See Walsh, 2015b for a scientifically-informed defence of organismal agency and a demonstration of agency's indispensable role in evolution.

One way to choose between various theories of metaphysical emergence is to test them against Kim's infamous exclusion argument (Kim, 1992, 1998, 1999). Kim's exclusion argument is widely and painstakingly discussed in the emergence literature, so here I do not spend many words describing it in detail.[22] Roughly speaking, Kim's exclusion argument runs as follows. Any emergent entity supervenes on some physical parts forming a physical base for that entity. This means that any changes in the emergent entity are caused by some changes in that physical base ($P \Rrightarrow E$). As the base is physical, any changes in the base should necessarily have a completely physical sufficient cause ($P \rightarrow P$). On these assumptions, one does not need any emergent causes to get a complete causal picture ($E \Rightarrow P$, $E \rightarrow E$, and $E \rightarrow P$) because the emergent causes can be re-described in terms of the physical causal interactions of the base ($P \rightarrow P$) and the supervenience causal interactions ($P \Rrightarrow E$) and, thus, get discarded. Because we can have a complete causal picture without the emergent causes, introducing these causes results in unacceptable causal overdetermination. It follows that emergent phenomena cannot have causal effects. In other words, we cannot have any of $E \Rightarrow P$, $E \rightarrow E$, and $E \rightarrow P$ types of causes. Kim concludes that emergent phenomena can only be epiphenomenal and, consequently, cannot be proper objects of science.

Kim's exclusion argument has been one of the most important challenges facing theories of metaphysical emergence, and different theorists have tried to address it in different ways (Gibb, 2019; Horgan, 1997; Kallestrup, 2006; Walsh, 2015a; Warfield & Crisp, 2001). Some theories of emergence provide better responses to Kim's exclusion argument than others, which is partly because of their background ontological assumptions. In my following discussions of various theories of emergence, I will use Kim's exclusion argument to evaluate these background ontologies. It will turn out that some ontological assumptions provide stronger bases against Kim's argument than others and are thus better suited for metaphysical emergence.

There are various structuralist theories of metaphysical emergence. I categorise these theories into two general approaches. The first approach, which I call *reductive structuralism,* uses a structural description of emergence to describe how emergent entities can be real and causally effective and, nonetheless, fully reducible to the parts. The second approach, which I call *non-reductive structuralism,* uses the concept of real structures to show how emergent entities are not reducible to the parts. These two categories come with their own ontological assumptions. I argue that only the second approach provides a good ontology for metaphysical emergence. Before going to that discussion, however, let me clarify one key term that I have repeatedly used in my discussion so far:

[22] For a careful and critical unpacking of the argument see Gibb, 2019.

distinctness. In the next subsection, I replace distinctness by indecomposability and provide what, I hope, are more precise versions of the metaphysical and causal claims.

4.2.3 Distinctness as indecomposability

Although defining emergence by causal and metaphysical distinctness is common in the literature, the precise meaning of "distinctness" in those definitions is vague and thus, the concept of emergence so illuminated, remains yet elusive.[23] Distinctness can be interpreted in many ways such as "not identical with", "separate", "independent from", and so forth, and each of these interpretations are, in turn, open to their own many understandings. The ambiguity runs so deep that I think many non-emergentists and emergentists alike could potentially agree that the whole is somehow "distinct" from the parts, each by their own understanding of distinctness. Therefore, it is imperative for emergentists who define emergence by causal and metaphysical distinctness to precisely clarify what they mean when they use the word. Here I suggest defining and replacing distinctness by indecomposability. This replacement results in the following versions of the metaphysical and the causal claims of metaphysical emergence:

> **The metaphysical claim**: There are systems that, as a whole, form real entities that are indecomposable to their constituent parts.
>
> **The causal claim:** The emergent entities have causal powers that are indecomposable to their constituent parts' causal powers.
>
> And the structuralist versions of the two claims would be:
>
> **The metaphysical claim (structuralist version)**: There are systems that, as a whole, form real structures that are indecomposable to their constituent parts.
>
> **The causal claim (structuralist version):** The emergent structures have causal powers that are indecomposable to their constituent parts' causal powers.

Indecomposability is not a new concept for emergentists. Emergence has long been associated with indecomposability, especially in the contexts where emergence is claimed to hinder mechanistic understanding (for examples, see Bechtel & Richardson, 2010, and Tabatabaei Ghomi, 2023). My suggestion here is taking this fundamental emergentist concept, and making it a criterion for emergence, rather than seeing it as merely an indicator or consequence of emergence. For indecomposability to play that definitory role we need to distinguish *epistemic* indecomposability

[23] I thank Neil Dewar and Denis Walsh for helpful comments and discussions on this point.

from *metaphysical* indecomposability as follows, because it is only the latter that can be the defining character of metaphysical emergence.

Indecomposability refers to the situation where decomposition is not possible. So, how we understand indecomposability, and consequently, how we understand the above formulations of the two claims of emergence, hinges on our understanding of decomposition. Decomposition can be epistemic or metaphysical. Epistemically, decomposition is a methodology for describing and understanding systems: we describe and thus understand a system by dissecting it into its individual constituent parts, and describing the function of each isolated part and its interactions with others. The resulting anatomical description of the system, theoretically, allows us to assemble individual parts and "recompose" the original systems (Bechtel & Richardson, 2010). Once we close a theoretical decomposition-recomposition cycle, so the method goes, we have completely understood the system. Biological investigation provides many good examples (see Chapter 2 for a more detailed discussion). For instance, biologists describe a cell by decomposing it into different organelles and describing the function of each organelle and its interactions with other organelles. The function of each organelle, in turn, is explained by describing its constituting parts and their interactions. Theoretically, this layered decompositional picture allows biologists to reconstruct the system from the bottom up by putting the parts together. The biological system is depicted as a jigsaw puzzle that can be taken apart into its pieces and put back together and so be explained and understood in terms of its pieces. Epistemic decomposition refers to this "divide and conquer" explanatory process that allows us to understand the whole in terms of the parts.

Metaphysical indecomposability, on the other hand, is not a method, but refers to a constitutional characteristic of a system or an entity. The nature of a metaphysically indecomposable system or entity is such that it cannot be decomposed into its parts. Such an entity can be monolithic and simple, but it can also be compound, and those compound indecomposable systems are the emergent ones. A simple entity, such as a hypothetical fundamental atomic (in the literal sense of the word) particle, is inherently indecomposable because it is simple. But as it does not have any parts, it satisfies neither the causal nor the metaphysical claims of emergence, and thus, it is not emergent. Emergence manifests when an entity is compound and yet, metaphysically indecomposable. An indecomposable compound entity rings paradoxical. How can we have a system that is constituted of parts but cannot be decomposed into those? This paradox is exactly what makes emergence such a resisting puzzle. But a correct understanding of causal and metaphysical indecomposability and examining the many examples of emergent systems around us, makes one appreciate that this paradox, as puzzling as it is, is not an unacceptable inconsistency.

A nice metaphor to describe the indecomposability of compound emergent systems is to say that they "break" into their parts, referring to the fact that they lose their entire metaphysical identity as well as their sort of causal powers by getting cut into pieces. Think about your own self as the closest example. Introspectively, we see ourselves as one unit even though we are compound systems constituted by many billions of parts. Should a daemon decompose us into our parts, we "break", meaning that our "I" would not exist anymore, and its metaphysical identity and causal powers get completely lost. The type of individualistic metaphysical identity and the sort of causal powers such as free will that we associate with our individual "I" are of a completely different nature compared to the metaphysical identity and the causal powers of our parts, and they get lost upon decomposition into parts. Decomposition "breaks" us.

The indecomposability of many other putatively emergent cases, however, is less clear. Take water for an example. Modern science has taught us that water is not a simple element but is a compound molecule constituted of hydrogen and oxygen, and a body of water, say, the stream of river Cam, is constituted of many of these molecules.[24] There are many features of a body of water such as its fluidity and transparency that are lost if we decompose water into individual water molecules. And in turn, there are many characteristics of those water molecules such as their molecular orbitals that are lost once we decompose each of them into its individual atoms. But is loss of fluidity or molecular orbitals enough to say that a body of water or a water molecule is metaphysically indecomposable? Opinions diverge here, and in order to find the correct answer and delineate the truly indecomposable, we need to specify the criteria for the type of characteristics the loss of which upon decomposition shows the indecomposability of a system.

Specifying these criteria is a hard intellectual challenge and I will not fully address it here. But let me entertain some options. One option is to ask that these characteristics should be intrinsic. Intrinsic properties are those that are determined solely by the internal constitution of the entity itself, and so are constant irrespective of the context in which the entity is situated (Walsh, 2015a). But this criterion seems to be too strong as it excludes some of the prime examples of indecomposable characteristics. For example, it is not obvious that a human self would manifest free will irrespective of the context. Another option is to ask that these characteristics should be non-relational. For instance, by this criterion, the loss of the relational character of "quenching thirst" upon decomposition does not show that water is indecomposable, while loss of the non-relational character of fluidity does. The good thing about this criterion is that it allows us to exclude tools and machines from the class of emergent entities. A hammer is not emergent even

[24] 3.34×10^{25} molecules per liter to be exact.

though it loses the character of hitting nails if we decompose it. The problem of this criterion, however, is that it is too weak and allows some non-relational and yet, obviously non-emergent properties such as having a certain volume or mass to pass.

Specifying the criteria for the type of characteristics the loss of which shows metaphysical indecomposability is an important and hard challenge, and I set it aside here for the future. But let me refer back to my metaphor of "breaking" upon decomposition and the prime example of our individual "I" that provide us with an intuitive appreciation of metaphysical indecomposability, which we can use as our "working" understanding of the concept for the time being.

A metaphysically indecomposable entity is epistemically indecomposable as well. If the nature of a system and its governing laws get lost in the decomposition process, we cannot understand that system by decomposition. Therefore, in cases where our attempts to understand a system by decomposition constantly fail, or in other words, when we face unresolvable epistemic indecomposability, one explanation for our failures could be that the system is metaphysically indecomposable. I developed this line of thought in Chapter 2. However, epistemic indecomposability can have reasons other than metaphysical indecomposability. For example, we may fail to come up with a decompositional description of a system because of our technical or epistemic limitations while the system itself is decomposable.

Failure of decomposition as an epistemic methodology, for whatever underlying reason, results in *epistemic* emergence. Epistemic emergence exists when a system *appears* to defy decomposition. This appearance of indecomposability can be transitory or permanent, depending on its underlying reasons. If these underlying reasons are resolvable, then epistemic emergence will be only a transitory state that fades away as the underlying reasons get resolved (Hempel & Oppenheim, 1948). For example, suppose biologists fail to decompose a system simply because they yet lack the necessary technology. In that case, the epistemic emergence of that biological system lasts only as long as the necessary technologies are developed. On the other hand, if the epistemic hurdle to decomposition of a system is insurmountable, then epistemic emergence would be a permanent feature of that system. One such insurmountable hurdle is the metaphysical indecomposability of the system itself, but this is not the only one. The insurmountable hurdle may be the upper bounds of our epistemic or technological abilities. For instance, decomposing the weather system into its elements may be beyond the upper bounds of our cognitive or technological abilities and so

weather might remain forever epistemically emergent, even though it might be metaphysically decomposable.[25]

Whatever its underlying reasons are, epistemic indecomposability is a claim about the failure of a specific method of explanation, namely, decomposition, and as such, it is different from metaphysical indecomposability which is a claim about certain characteristics of a system. Those interested in epistemic accounts of emergence can use indecomposability in its epistemic sense. Here, however, we are discussing metaphysical emergence and therefore, the references to indecomposability in the metaphysical and causal claims should be interpreted in the metaphysical sense.

I leave this discussion at this point and continue to use the term "distinct", as a common term of art in the emergence literature. However, I hope that the discussions of this section have clarified what we talk about when we talk about metaphysical and causal distinctness.

4.3 REDUCTIVE STRUCTURALISM

Reductive structuralism refers to all those theories of metaphysical emergence that describe emergent entities as structures to maintain two, in my opinion, contradictory commitments at the same time: a commitment to the two claims of metaphysical emergence, and a commitment to a reductive ontology in which everything is ultimately reducible to physical parts. The commitment to reductive ontology gets a higher priority and the theorists in this category drop or weaken their commitment to the claims of metaphysical emergence should these claims clash with their ontological commitment.

I discuss two theories in this category, Deacon's and Wimsatt's, that reflect different flavours of reductive structuralism. Deacon's theory is a weak version of metaphysical emergence that only weakly and partly commits to the two claims of metaphysical emergence. Wimsatt's theory endorses the two claims of metaphysical emergence, but provides weaker interpretations of those claims compared to what we will see in *non-reductive structuralism*. Both theories are committed to a reductive ontology, and we will see that this ontological commitment makes these theories vulnerable to important objections. The general conclusion that I hope to draw from the discussions of this section is that a metaphysical emergentist has no way but to give up commitment to the ontology of generative atomism and start looking for alternative ontological

[25] I put indecomposability due to any practical hurdle under epistemic emergence, although some practical hurdles might not be, strictly speaking, epistemic. I prefer to accept this slight conceptual imprecision to avoid introducing a third class of *practical* emergence on top of epistemic and metaphysical emergence.

options. The efforts of reductive structuralists to mix the water and oil of generative atomism and metaphysical emergence are all futile. Then, in the subsequent sections, I suggest the outlines of viable alternative ontological options by proposing the list of minimum ontological necessities for metaphysical emergence.

4.3.1 Deacon's theory of emergence

Deacon describes emergence in terms of structural constraints (Deacon, 2006; Deacon & Cashman, 2012). He classifies his theory as an instance of metaphysical emergence (Deacon, 2003), even though he is not fully committed to the two claims of metaphysical emergence. [26] In particular, Deacon acknowledges emergent causal powers but denies that emergent entities are new *things*. Although he acknowledges emergent causes, he emphasises that the emergent causal powers are not on par with the fundamental causes of the fundamental particles. Rather, he suggests that emergent causes arise from structural constraints imposed on the fundamental causes, and thus, he calls those *configurational causes* (Deacon, 2007). According to him, emergent configurational causes appear only because matter and energy, i.e. the carriers of fundamental causal powers, are arranged in certain structures and, thus, act within certain constraints. These structural constraints exert patterns on how the fundamental causes interact (Deacon, 2006) and allow those fundamental causes to perform new functions which we associate with the system as a whole. As he and Cashman (2012: 204) put it, "[a]mazing new properties have been, and are being, emerged, and there is nothing new being added. There is no thing new. No new laws. What is 'new' and 'more' are new modes of not being, new forms of constraint."

But the concept of emergent causes, defined as functions resulting from the arrangement of matter and energy in certain ways, can potentially trivialise the idea of emergence. My fence has the function of forming its unique shade because its constituent matter and energy are constrained in the particular arrangement of its panels. Does my fence have emergent configurational causal powers? And if it does, are its configurational powers of the same nature as the configurational powers of my brain? To avoid triviality, Deacon needs to qualify somehow the structures and the constraints that result in emergent configurational causes, and indeed, he suggests such a qualification. Deacon (2003) claims structures that result in emergent phenomena show *topological reinforcement* or *amplification in pattern formation*, where amplification is the recurrent superimposition of the same structure. Thus, for Deacon, only a particular type of arrangement of matter and energy, and not any trivial such arrangement, results in emergent phenomena.

[26] Deacon introduces three levels of emergence, each with its own unique characters. For more details about his theory see Deacon, 2006, and 2007. Here, I give only a broad overview of his account of emergence.

He gives a few examples of such amplificatory structures, but none make a convincing case for his theory of emergence. One of his examples is the amplificatory process that results in the astonishing forms of snowflakes. This is not a good example for Deacon because the resulting formations do not seem to perform any particular functions. Another one of his examples that is supposedly functional is a process he calls "autocatalysis." Autocatalysis happens when a catalyst molecule can catalyse its own formation. If the molecule has some evolutionary value, autocatalysis can result in the proliferation of the autocatalysing molecule, or in Deacon's terms, amplification of its pattern. Deacon brings other examples as well. But the problem with his examples is that they seem to be exceptions rather than rules. Even if we find particular examples in biology that more or less conform to the autocatalysis process as described above, they seem to be particular cases rather than representatives of all emergent biological functions. Many amplificatory patterns are indeed identified in nature (Kappraff, 2001; Mandelbrot, 1982), but many of those do not result in any particular function, and many emergent functions do not seem to have resulted from such pattern amplifications.

A deeper problem with Deacon's theory is that he does not point to any fundamental difference between amplificatory structures and any other structures to justify separating the former, calling it emergent, and associating it with configurational causal powers. Non-amplificatory structures also constrain fundamental causes and can potentially make those causes perform functions they would not be able to perform outside of those structures. So why should we give the functions resulting from amplificatory processes the special title of "configurational causes" and treat them differently from those resulting from non-amplificatory processes?

In summary, Deacon's theory does not commit to the metaphysical claim of metaphysical emergence and fails to support the causal one. His theory fails to justifiably separate the structures he associates with emergence from any other structures. And he fails to provide a sufficiently large and diverse set of examples to justify his theoretical claims. All these, I believe, partly result from his ontological commitments. He seems to be strictly committed to an ontology in which everything is reducible to fundamental particles and their causal powers. He strictly rejects that emergence can have conflicts with this reductionist ontology. He writes:

> The concept of emergence probably has gained its worst reputation when it has been used in a primarily negative sense—to point to something missing in reductionistic explanations. … In this use, the concept of emergence is a place-marker intended to indicate points where standard reductionistic accounts fail or seem incompletely to explain apparent discontinuities in properties exhibited at different levels of physical scale. This negative usage has

> unfortunately led many more orthodox thinkers to suspect that there is no underlying phenomenon to be described, only a vague suspicion due to incomplete analysis. (Deacon, 2003:274)

Siding with these "more orthodox thinkers", he aims to avoid this so-called negative concept of emergence and remain fully committed to reductionism. But his full commitment to the reductionist ontology makes Deacon's world too small to accommodate metaphysically real emergent entities or causes. Ultimately, his reductionistic ontology is occupied by atoms and atoms only. His structural characterisations of emergence cannot go beyond descriptions of some particular arrangements of these atoms. None of his structural characterisations of emergence can break out of this ontological cage. And this is why, as he puts it in the quote referred to at the beginning of this sub-section, "no thing new", and "no new laws" can come out of his theory, and the result of his view is restricted to "modes of not being." "Not being" is not enough for metaphysical emergence.

4.3.2 Wimsatt's theory of emergence

Let us shift to Wimsatt's (2000) theory of emergence as a highly influential reductive structuralist account of metaphysical emergence. Wimsatt starts by the observation that scientists usually use the concept of emergence in ways that are compatible with reductionism, while philosophers' descriptions of the concept are usually anti-reductionistic. He aims at developing a philosophical concept of emergence that is scientifically motivated, and so it is compatible with reductionism. At the same time, Wimsatt wants this notion of emergence to resist what he calls *vulgar reduction* or *nothing-but-ism*. He wants his theory to establish the emergent phenomena's ineliminable importance and ontological reality. Clearly, he sees emergent entities as real occupants of the ontology of our world. He writes:

> These are higher-level ontological features, Organizational *Baupläne*, related to the things that people usually talk about under the topic of ontology (things like objects, properties, events, capacities, and propensities) as paragraphs are to words and phonemes or morphemes. But they are there nonetheless; it is only our concern with the little things, motivated by foundationalist or reductionist concerns, which has deflected our attention from them. (Wimsatt, 1994:208) (emphasis in the original)

So, he is fully committed to both reductionism and the claims of metaphysical emergence.

Wimsatt's recognition of emergent entities as real occupants of the world's ontology is an important difference between his and Deacon's theories of emergence. Nonetheless, like Deacon,

Wimsatt is committed to the reductionistic picture of generative atomism and claims that these emergent entities and their associated properties result from some particular structural arrangements of the parts. Therefore, in principle, they are reducible to the parts and their arrangements. But, similar to Deacon, Wimsatt qualifies the type of arrangements that result in emergence. According to him, emergence happens when parts are arranged in non-*aggregative* structures. Non-aggregativity is best understood by contrasting it with aggregativity. Completely aggregative systemic properties satisfy four conditions. A system may satisfy only some of these properties or satisfy them only to a certain degree and, therefore, be only partly aggregative or, put the other way, partly non-aggregative.

The first condition is *inter-substitution.* A systemic property is inter-substitutive if it is invariant under re-arrangements of parts or substitution of parts with other parts from an equivalent class. For instance, consider the mass of the air in a room. This mass does not change by the movements of air molecules around the room or by replacing all the oxygen atoms inside the room with the same number of oxygen atoms collected from outside the room. The second condition is *size scaling* which means that the addition or subtraction of parts affects the systemic property only quantitatively and not qualitatively. For example, if we remove half of the atoms of the air in the room, the mass gets quantitatively reduced in half but remains qualitatively unchanged. The third condition is *re-aggregability*, which means that the systemic property can be regained upon decomposition and recomposition of the system. For example, if you separate every atom of the air in the room and then mix them again, you get the same mass. And the last and fourth condition is *linearity*, which means there should be no cooperative or inhibitory feedback or feedforward interaction loops among the parts that affect the systemic property. Looking again at the mass of the air in a room, the property results from a simple addition function and no interaction loops affect its value.

The mass of the air in a room perfectly satisfies all four conditions, and therefore, it is completely aggregative. In other words, it is completely non-emergent. But many systemic properties do not fully satisfy these conditions. Wimsatt suggests that emergence happens when, and as much as, a systemic property does not satisfy the conditions of aggregativity. A complex biological function, for instance, does not fully satisfy any of these conditions. Re-arrangement and replacement of parts in biological systems can result in significant functional changes (Lu et al., 2009). You cannot cut a biological system in half and expect to get the same function only quantitatively cut in half. Attempts to decompose and recompose biological systems often fail (Kwok, 2010; Tabatabaei Ghomi, 2023). And biological systems are replete with non-linear

interactions (Mazzocchi, 2008). Thus, by Wimsatt's definition, biological systems are non-aggregative and emergent.

Non-aggregativity partly depends on the type of property we are dealing with and partly on how the parts are arranged. So, Wimsatt's emergence is partly structural. Wimsatt emphasises that non-aggregative emergent entities are objective facts of nature, can be reliably identified by empirical methods, and form an ineliminable part of our science (Wimsatt, 1994). This endorsement of the metaphysical claim is an advantage of Wimsatt's account over Deacon's. Another advantage of his account is that his four conditions for emergence are more comprehensive and inclusive compared to Deacon's criteria of amplification.

Nonetheless, his ontology is still under full governance of reductionism. Surveying different putative cases of emergence, he writes:

> Many cases were classically considered as involving emergence – cases motivating claims that "the whole is more than the sum of the parts." – like an electronic oscillator circuit. There's nothing antireductionistic, mysterious, or inexplicable about being an oscillator. You hook up an inductance, a capacitor, and a resistor in the right way with a voltage source. The system oscillates although none of its parts in isolation do so. Moreover, it is how these disparate parts are strung together that makes them an oscillator. (Thus, an oscillator must contain a closed circuit.) A deductive theory relates properties of the parts to the frequency and amplitude of the oscillator; reductionistic even by the strong conditions of the formal model of reduction, and also by the weaker characterisation given here. This is intuitively a case of emergence, though it can't be if we tie emergence to non-reduceability.
>
> More generally, *emergence of a system property relative to the properties of the parts of that system indicates its dependence on their mode of organisation.* It thus presupposes the system's decomposition into parts and their properties, and its dependence is explicated via a mechanistic explanation. … We need an analysis of emergence consistent with reductionism. (Wimsatt, 2000:271)(emphasis in the original)

Because of its full-blooded commitment to reductionistic ontology, contrary to Wimsatt's claims, his theory fails to support the causal claim of metaphysical emergence. At the end of the day, these are the non-aggregative interactions of the causal powers of the parts that result in the emergent entities and their functions. Even if the emergent entities and functions are methodologically required for scientific modelling, metaphysically speaking, they are powered by the causal interactions of the parts. Therefore, although Wimsatt wants to resist nothing-but-ism,

causally speaking, his emergent causes seem to be nothing but fundamental causes arranged in certain non-aggregative ways. This makes Wimsatt's account vulnerable to Kim's exclusion argument. Why do we need emergent causes if we can explain everything in terms of the causal powers of the parts arranged in a particular type of structure? Following Kim's conclusions, the emergent entities Wimsatt recognises as real occupants of the world can only be epiphenomenal.

I believe any other theory of reductive structuralism fails in the same ways that Deacon's and Wimsatt's theories fail because, as we saw, their failures have their roots in their commitment to the reductionistic ontology. On this ontological background, parts have metaphysical and causal primacy over emergent entities. Any structural characterisation of emergence that is restrained by this ontological commitment can, at most, describe some interesting features of emergent entities and cannot establish their causal and metaphysical independence. But structural characterisations can do much more in non-reductive structural theories that do not assume the reductionistic ontology. Let us now turn to these theories.

4.4 NON-REDUCTIVE STRUCTURALISM

By non-reductive structuralism, I refer to the theories of emergence that use structural descriptions to show how the emergent level is irreducible to the lower levels. We will see that these accounts are not committed to the ontology of generative atomism and are, therefore, far more successful in maintaining the claims of metaphysical emergence and responding to Kim's argument. I discuss two theories in this category. First, I discuss Walsh's (2012, 2015b) reflexive emergence as a representative of a prevalent approach in scientific emergentism that suggests that the emergent causes exert their effects by affecting how the parts behave within the system (Gillett, 2019).[27] Then I discuss Santos' (2015) view of emergence as a more radical form of structuralism that describes both the emergent and the lower levels as structures and, therefore, inherently of the same nature. From these accounts, we will learn ontological lessons that will inform the construction of our minimum set of necessary ontological commitments for metaphysical emergence.

4.4.1 Walsh's reflexive emergence

When writing about emergence, Walsh's primary concern is biological teleology and the agency of organisms as a higher-level cause (Walsh, 2012, 2015b). His resulting theory of emergence, however, is general and not limited to teleology. According to Walsh, higher-level systemic causes

[27] Walsh does not give his account of emergence any particular names. The title of "reflexive emergence" is mine.

such as agency exert their effects via what he calls "reflexive downward regulations" (Walsh, 2012). Reflexive downward regulations are reciprocal regulatory and causal interactions that exist between an emergent system as a whole and its parts. The parts, their powers, and their relations form the system, and in turn, the system as a whole modulates the parts, their powers, and their relations. For example, while it is the amino acids and their causal interactions and properties that make a protein what it is, the effects of the amino acids and subsequently the expected outcome of their mutations is reciprocally under the causal influence of the protein as a whole system. The systemic downward regulatory effects of a protein system on its constituent amino acids can be so salient that a mutation that contributes favourably to some systemic property of some protein A (e.g. catalysis, thermostability etc.) might turn out to be unfavourable when introduced in the sequence of a slightly different protein B (Starr & Thornton, 2016; Yoo et al., 2020).[28] Systemic causal effects, Walsh suggests, work via these kinds of regulations that systems exert on the causal properties of their parts (Walsh, 2014, 2015a).

Walsh is minimalistic in the type of emergent causation he underwrites for emergent wholes. His reflexive downward regulation only consists of $E \Rightarrow P$ type of causes, and it seems that he sees $E \longrightarrow E$ and $E \longrightarrow P$ type of causal relations to hold via $E \Rightarrow P$ causations of reflexive downward regulations via a chain like $E_1 \Longrightarrow P_1 \longrightarrow P_2 \Rrightarrow E_2$ (see Figure 4.2 and section 4.2.2). This is enough to allow some causal effect for emergent entities and endorse the causal claim of metaphysical emergence. The question, however, is whether this type of minimalistic emergent causation is immune to Kim's argument. Walsh thinks reflexive downward regulations are immune to Kim's argument because the higher-level causes are distinct from the causal powers of the parts and form inalienable components of a complete causal picture (Walsh, 2015a). This solution, however, works only if three conditions are satisfied. Walsh mentions two of these conditions, and I add the third one.

The first condition is that the causal capacities of the parts should not be, in Walsh's terms, "intrinsic". He defines intrinsic properties as the properties that an entity has irrespective of the context in which it is situated. Intrinsic properties are determined solely by the internal constitution of the entity itself (Walsh, 2015a). Walsh contends that emergence happens where the parts' causal powers are not intrinsic, and the system as a whole causally modulates the non-intrinsic causal powers of its parts. I think we can understand this condition in a relaxed way saying that to have

[28] A phenomenon called *epistasis*.

reflexive emergence we need at least *some* of the casual capacities of *some* of the parts to be non-intrinsic. This is enough to allow downward regulatory effects.

The second condition is that the non-intrinsic causal powers of the parts should be regulable by higher-level causal effects. If the parts' non-intrinsic causal powers are regulable only by causal interactions with other parts, the higher-level causal powers will become redundant. This malleability by higher-level causes is a key difference between non-reductive structuralist accounts of emergence such as Walsh's, and reductive ones such as Deacon's. Reductive structuralists like Deacon do not necessarily disagree that parts' casual capacities could be non-intrinsic and malleable. They are even fine with the system putting some sort of constraint on how the parts' causal powers act. But eventually, these are only the parts that are causally effective and can affect each other, and higher level causes such as agency play no real causal role. Walsh's reflexive emergence, however, endorses systemic causal powers that can manipulate and regulate parts' behaviours. Consequently, his account allows an organism to have agency.

On top of the above two conditions of reflexive emergence set by Walsh, I suggest that we also need a third condition for reflexive emergence to resist Kim's argument effectively. We need the downward regulatory effects to be possible at every level, and all the way down to the most fundamental level. Otherwise, one can reconstruct Kim's argument at the lowest level where all causal effects are intrinsic and the emergent causal effects cannot affect them. Let us call that level L_i. At L_i, the intrinsic causes do not get affected by downward regulatory effects and we cannot have $E \Rightarrow P$. In reflexive emergence, downward regulatory effects ($E \Rightarrow P$) are the only way emergent phenomena can be causally effective and unmediated, genuine $E \longrightarrow E$ and $E \longrightarrow P$ are not allowed. Therefore, in absence of $E \Rightarrow P$ at L_i, the emergent entities will turn out to be epiphenomenal. Just group all emergent causal effects at L_i and above and call them E, group all lower lever level causes at L_i and below and call them P, and construct Kim's argument between E and P.

The conclusion is if downward regulatory effects ($E \Rightarrow P$) are the only way emergent phenomena can be causally effective, we can have non-epiphenomenal emergent entities only up to the lowest level below which we do not have any levels where causal effects are exclusively intrinsic. This is equivalent to saying that if downward regulatory effects ($E \Rightarrow P$) are the only way emergent phenomena can be causally effective, then downward regulatory effects of any non-epiphenomenal emergent entity should be able to percolate down to all of the causal levels beneath it. Those downward regulatory effects should not hit the barrier of any L_i-like level. In other

words, if emergent causal effects are exclusively downward regulatory effects ($E \Rightarrow P$), it is not enough to have downward regulation only at the higher levels such as biology and chemistry. Regulations should be able to go all the way down to the most fundamental physical level. Let us call this third condition *the condition of deep percolation*.

I am not aware of any instance where Walsh explicitly discusses the idea behind the condition of deep percolation. Still, I do not think that he would necessarily object to the idea, as at one point, he argues that even the seemingly rock-solid fundamental property of mass is not intrinsic:

> The mass of a particle appears to be an 'emergent' property (Smolin, 2013). It arises out of the interaction of a particle with the Higgs field. Mass is 'conferred' upon a particle by its relation to something else. So, causal powers, even if they are invariant across a wide range of contexts, can be relational, properties of things. (Walsh, 2015a:221)

However, it is important to underline that allowing the downward regulatory effects of higher levels to percolate all the way down to the most fundamental level means allowing biological, psychological, and social phenomena to potentially regulate the most fundamental laws of physics. Walsh thinks "(e)mergence ought to hold no fears for contemporary science" (Walsh, 2015a: 218). Biology regulating physics, however, will surely scare many contemporary scientists. But this is an inevitable price Walsh must pay for a solid reflexive emergence.

This ontological picture in which biology can regulate physics sparks the idea that the metaphysical nature of the higher-level entities such as the biological ones, and the lower-level entities of fundamental physics might not be that different after all. We will see an extreme version of this idea in Santos' relational emergence discussed next. By claiming that lower-level and higher level entities are all of the same nature, Santos's theory expands the causal options of metaphysical emergence and allows $E \longrightarrow E$ and $E \longrightarrow P$ causal relations on top of $E \Rightarrow P$.

4.4.2 Santos' relational emergence

Santos' emergence theory is based on what he calls a "relational ontology" (Santos, 2015). In this ontology, reality is nothing but a web of interrelated relata. Entities are relata that can be systematically distinguished by their set of stable internal or external relations. Therefore, "the existence-conditions, the identity and the causal behaviour of any entity must always be conceived and explained as constructed and transformed by the interplay of its intrinsic and extrinsic relational processes." (Santos, 2015:439). These relational existence, identity, and transformation conditions apply not only to the emergent entities, but also to the parts of those emergent entities, and the parts of the parts, all the way down to the most fundamental parts.

The existence and identities of entities such defined are fluid under the influence of the entities' internal and external interaction networks. Emergence in this ontology refers to the process where an entity undergoes qualitative changes under the influence of its internal or external interactions, or when a group of entities interact and form a new holistic entity with new properties. The fluidity of existence and identity applies to the fundamental entities as well. The fundamental entities are also defined as relata, and their properties can change under the influence of their interactions. These interactions need not be at the fundamental level. Higher-level entities can interact with and thus affect the fundamental entities too.

Describing both the emergent and the lower levels as structures and, therefore, inherently of the same nature is in contrast with the theories of emergence described above in which we find two realms, or modes of existence: the realm of physical parts and their properties, and the realm of emergent entities and their properties. The first realm is material, and the second is structural. In Santos' view, there is only one realm, the realm of structures, and the parts and the emergent entities both belong to this same single realm. Everything, even the most fundamental elements, is structure.

As all the real entities in Santos' ontology have a structural nature, it is straightforward to claim that emergent structural entities are *real*; they are as real as anything can be. Biological species are as real as atoms. Moreover, as there is no difference between the nature of entities at different levels (all are structures), there is also no difference between the nature of their causal interactions. $P \rightarrow P$, $E \rightarrow E$, $P \rightarrow E$, and $E \rightarrow P$ all belong to the same type of causal relations and are all equally acceptable. Santos' relational ontology, therefore, easily supports both the causal and the metaphysical claims of metaphysical emergence and draws an ontological picture in which both claims make sense. Also, the possibility of all sorts of the emergent causal relations in Santos' relational ontology as well as the fact that emergent phenomena and their properties result from the totality of their multi-level web of interactions and not merely the lower-level interactions among their parts mean that Kim's argument completely fails in this ontology.

At first sight, Santos' relational ontology may seem too scientifically unorthodox and philosophically imaginative. Still, it draws support from the scientific evidence and philosophical arguments supporting so-called eliminative radical ontic structural realism. Eliminative radical ontic structural realism is a sub-type of ontic structural realism. In general, ontic structural realists hold that entities ontologically depend on structures. *Moderate* ontic structural realists hold that entities and structures reciprocally depend on each other, while *radical* ontic structural realists hold that entities ontologically depend on structures but not vice versa. Finally, *eliminative* radical ontic

structural realists hold that entities do not exist in the real sense, and all that really exists are structures (Dewar, 2019). Santos' relational ontology belongs to this last class. There is a rich literature in support of various forms of ontic structural realism (Dasgupta, 2011; Dewar, 2019; Esfeld, 2004; French & Ladyman, 2010; French & Redhead, 1988; Frigg & Votsis, 2011; Kantorovich, 2003; Ladyman, 1998; Ladyman & Ross, 2007; Poincaré, 1952; Worrall, 1989). I do not have the space here to discuss this literature. But those who are moved by the philosophical arguments and the scientific evidence cited in that literature would find Santos' relational ontology much easier to swallow (Cordovil et al., 2022).

Santos' ontology, however, is an overkill, and a more moderate relational ontology suffices to support the two claims of metaphysical emergence and address Kim's argument. Santos claims that all entities are of relational nature, and all of their causal and non-causal properties depend on their web of interactions. Such a strong view is sufficient, but not necessary for metaphysical emergence. It is enough to say that there exist *some* real entities of structural nature, and that at least *some* aspects of their identities are determined by their multi-level web of interactions. This is enough to allow metaphysically distinct structural emergent entities. Also, to uphold the causal claim and ward off Kim's argument, we do not need to accept all different types of emergent causations ($E \rightarrow E, E \rightarrow P, E \Rightarrow P$), and we do not need to claim that these are the emergent causes that exclusively govern the world. It suffices to accept only some type of emergent causation in some cases. In other words, we do not need eliminative radical ontic structural realism to have metaphysical emergence. Even some version of moderate ontic structural realism is enough.[29] One may subscribe to eliminative radical ontic structural realism for any reason. But this subscription is not mandatory for upholding metaphysical emergence.

[29] Traditionally, ontic structural realism is not widely taken up as an ontological background for theorizing about metaphysical emergence. One major reason behind this has been the anti-emergentist declarations and arguments of some major proponents of ontic structural realism that resulted in the wrong assumption that ontic structural realism is incompatible with metaphysical emergence. However, by careful unpacking of these anti-emergentist claims and statements, Cordovil et al (2022) show that the anti-emergentist views expressed by proponents of ontic structural realism are independent of their arguments in support of ontic structural realism and therefore, ontic structural realism and metaphysical emergence are not necessarily incompatible. They correctly point out that ontic structural realism and the arguments in its support can be used as an ontological setup for theorizing about metaphysical emergence. In fact, some supporters of ontic structural realism such as French (2014) and Kinacid (2008) have proposed extending the ideas of ontic structural realism developed in physics to the context of special sciences such biology and social sciences. Such extensions can be taken as attempts to explain emergent phenomena from the perspective of ontic structural realism.

4.5 THE MINIMUM SET OF NECESSARY ONTOLOGICAL COMMITMENTS FOR STRUCTURAL METAPHYSICAL EMERGENCE

Putting all the lessons that we learnt through our analysis of different types of structural metaphysical emergentism together, we come to the following minimum set of necessary ontological commitments for structural metaphysical emergence:

A) Structural realism: structural entities can be real.
B) Structural causation: at least some real structural entities can have irreducible causal effects on some other structural or non-structural entities at their own level or levels below them.
C) The condition of downward percolation: If the only type of irreducible causal effect that emergent structural entities can exert is $E \Rightarrow P$, then the causal effects of some real structural entities should be able to percolate all the way down to the lowest level beneath them.

I want to emphasise that this list is minimalistic and, as such, is compatible with some more expanded structuralist ontologies such as Santos', or even the "bottomless" structural ontologies that deny the very existence of a most fundamental level and posit an infinite regress of structures instead (Schaffer, 2003). These more radical ontologies work well for metaphysical emergence, and an emergentist might choose to subscribe to those ontologies for whatever reason. But she does not have to. Accepting the above minimum set of necessary ontological claims is enough to develop substantive and robust accounts of metaphysical emergence. However, this set is the necessary minimum, and we cannot get metaphysical emergence with anything ontologically cheaper than this minimal set.

Now let me expand a bit on Schaffer's bottomless ontology as an example of an ontology that satisfies the minimum necessary commitments and provides a very good ontological choice for metaphysical emergence.

4.5.1 An example of a suitable ontology for metaphysical emergence: Schaffer's bottomless ontology

Schaffer rejects that there is a fundamental level. Instead, he suggests that the world is a web of ever deeper and deeper structures. He starts by rejecting that an atomic most fundamental level is a logical necessity by showing that an atom-less metaphysics is conceivable, logically consistent, and physically serious. He concludes that, from the armchair, a world with a fundamental atomic

level and a world not having such a fundamental level are equally acceptable, and the judgement as to which one is the correct ontological description of our world is a matter of empirical investigation.

He then claims that, contra common belief, empirical evidence does not support the idea of our world being built on an atomic fundamental base. First, he contends that science is not close to discovering the most fundamental theory of physics, and a quick look at the history of science is enough to warrant dismissing all the brazen claims of the scientists who claim to be close to discovering "the" most fundamental theory. He suggests that such claims are useful myths serving practical purposes of the scientists who make them. Moreover, the scientists who make those claims disagree on the content of the most fundamental theory and so, from the perspective of us, the bystanders, they defeat each other's claims. Besides, some of those suggested fundamental theories such as quantum field theory are non-atomistic and this shows that even if one day scientists do find the most fundamental theory, that theory would not necessarily be about atoms.

Schaffer pushes further and claims that, in contrast to the prevalent atomistic view, the history of science provides a meta-inductive argument that the world is comprised of ever deeper and deeper structures. He writes:

> Indeed, the history of science is a history of finding ever-deeper structure. We have gone from "the elements" to "the atoms" (etymology is revealing), to the subatomic electrons, protons, and neutrons, to the zoo of "elementary particles", to thinking that the hadrons are built out of quarks, and now we are sometimes promised that these entities are really strings, while some hypothesize that the quarks are built out of preons (in order to explain why quarks come in families). Should one not expect the future to be like the past? (Schaffer, 2003:503)

In this ontology, all levels are somewhere in the middle of an infinite chain of structural levels, and therefore, there is no fundamental difference between levels. Take any level you like. That level will have infinitely many levels below it. So, effectively, all of the levels are at the same level. As Schaffer puts it, everything is "macro":

> The most striking feature of an infinite descent is that no level is special. Infinite descent yields an egalitarian ontological attitude which is at home in the macro-world precisely because everything is macro. Mesons, molecules minds, and mountains are in every sense ontologically equal. Because there can be no privileged locus for the causal powers, and because they must be somewhere, they are everywhere. So infinite descent yields an egalitarian metaphysic which dignifies and empowers the whole of nature (Schaffer, 2003: 512-513)

Schaffer only refers to infinitely many lower and lower structures. But what about infinitely many higher and higher structures? As far as the present discussion is concerned, it does not matter whether there are infinitely many higher structures, or the universes is capped with one highest "Whole" structure. If there is a Whole, then everything will be macro because, as noted, for any given level there are infinitely many levels below it. If there is no Whole and there are infinitely many higher structures, then everything will be *meso*, sandwiched between two infinities. Macro or meso, either way all levels would belong to the same cast.

This egalitarian bottomless ontology satisfies the minimum set of ontological commitments necessary for metaphysical emergence. But what is particularly nice about this ontology is that it provides an explanation for why these commitments hold. Here is the explanation. As everything is structural, and everything is at the same macro or meso level, entities at any level can be equally real (A is satisfied). And because there is no difference between the nature of the levels (all of them are macro or meso), all six different types of causal relationships ($P \Rrightarrow E$, $E \Rightarrow P$, $P \rightarrow P$, $E \rightarrow E$, $P \rightarrow E$, and $E \rightarrow P$) are potentially allowed, (B is satisfied), and there is no problem in causal effects of higher levels percolating deep down (C is satisfied).

4.6 CONCLUSIONS

The literature on metaphysical emergence is usually about the arguments for or against the claims of metaphysical emergence. I turned this common practice on its head. I assumed the claims of metaphysical emergence and asked how the world should be in order to allow them. I first argued for a specific yet widely encompassing version of structural metaphysical emergence. I then critically surveyed some structuralist theories of emergence to find the minimum ontological commitments needed to uphold the metaphysical and causal claims of metaphysical emergence and provide a robust response to Kim's argument. I concluded that structuralist metaphysical emergentists need to commit to at least three ontological claims, (A) structural realism, (B) structural causation, and (C) the condition of downward percolation.

These three ontological commitments are minimalistic. Nonetheless, they draw an unorthodox ontological picture of the world that is far away from the ontology of generative atomism commonly associated with the modern scientific worldview. This might be unsettling for those who are committed to the prevalent reductionistic ontology and at same time, have a warm spot in their heart for metaphysical emergence. Indeed, metaphysical emergentists face a formidable challenge to argue in support of the unorthodox ontology against the reductionist one. But let me motivate my reader to side with this unorthodox view of the world by emphasising its huge

emancipatory power: it frees the macro-level entities, objects of special sciences, and mental phenomena from the causal and metaphysical chains of fundamental physics. Consequently, it allows us to respond to longstanding conundrums with simple and intuitive answers. Do chiffchaffs and cherry trees really exist? Do they really cause chirps and cherries? Do chiffchaffs chirping on cherry trees really infuse us with something called joy, which in turn, causes us to murmur a happy song? The ontology of generative atomism makes us wonder. The alternative nonreductive structuralist ontology allows us to give the most intuitive answer: yes, they do!

Conclusions

5.1 RETROSPECT

In Chapter 1, I showed that computational irreducibility fails to fulfil its assigned philosophical roles and thus, took ontological weak emergence off the table. In Chapter 2, I concluded that irrationality suggests indecomposability and indecomposability implies emergence. In Chapter 3, I showed that half-hearted emergentism is expendable, and wholehearted emergentism is expensive. And finally, in Chapter 4, I constructed the minimum set of necessary ontological commitments for structural metaphysical emergence.

What does this all mean for chiffchaffs and cherries? Chapter 1 shows us that we cannot explain away chiffchaffs and cherries as epistemic side effects of some mathematical constraint. Chapter 2 shows us that the practice of science provides evidence that cherries, and not merely their atoms, are real occupants of the world and also, it is the chiffchaffs, and not merely their constituting atoms, that make the chirps. Chapter 3 shows that we should either take cherries as real entities distinct from their atoms and chiffchaffs to be the genuine sources of their chirps, or reject these claims altogether. A half-hearted middle ground is not tenable. And finally, Chapter 4 gives a minimalistic description of a world in which one can wholeheartedly embrace the reality of cherries and the chirping powers of chiffchaffs.

5.2 PROSPECT

What else can be said and done? Here I suggest two possible lines for future research. The first looks into the ramifications of metaphysical emergence for decision-making. The second is a proposal to understand beauty as a metaphysically emergent phenomenon. [30]

5.2.1 Metaphysical emergence and the precautionary principle

What are the practical consequences of accepting metaphysical emergence for policy decisions about new technologies? One of the main guiding principles for such decisions has been the so-called precautionary principle (Ahteensuu, 2007; Bodansky, 1991; Resnik, 2003). There are several formulations of the precautionary principle (Sandin, 2004; Steel, 2015). But the general idea behind

[30] There is also another important line for future research. For the philosophy of emergence to bear its best fruits, we need to diversify its landscape, and one way to do that is to bring ideas from non-Western philosophical traditions. There have been some movements in this direction. For example, Ganeri (2011) draws on emergentist ideas in Indian philosophy to respond to Kim's argument, and Shariati and Khoshnevis (2020) compare the ideas of Sadr al-Din Shirazi, a Persian philosopher of the 16th and 17th centuries, with those of William Hasker in the context of the philosophy of mind. These are good first steps, but much more must be done.

the principle is that in cases where a decision might have large adverse environmental or human health consequences, one should err on the side of caution even if one does not have sufficient evidence to believe that those consequences would occur. The precautionary principle has been subject to numerous criticisms, and even those who accept it agree that it should not be the only principle guiding one's decision (Bodansky, 1991; Bognar, 2011; Gardiner, 2006; Peterson, 2007; Sunstein, 2005; VanderZwaag, 2002). So, it is important to assess its validity and its weight in various cases of decision-making and from different angles. One angle to look at the precautionary principle is through the perspective of a metaphysical emergentist.

Metaphysical emergence has two opposing consequences for the precautionary principle, depending on how we see the matter. On the one hand, a metaphysical emergentist who asserts the possibility of emerging new entities with novel causal powers beyond those considered in the initial risk assessments and denies that emergent systems can be mechanistically understood has strong reasons to give considerable weight to the precautionary principle, especially in biological and environmental contexts that provide fertile grounds for emergence. On the other hand, a metaphysical emergentist who asserts that complex systems result in emergent entities with stable regularities opens the door for reliable and accessible risk assessment of complex systems on the emergent level. The systems so complex that they cannot be risk assessed based on a mechanistic understanding at the level of their parts might lend themselves to reliable risk assessment based on the rules and regularities of the emergent level. That higher level risk assessment might show that the risks are low and make a case against applying the precautionary principle. A metaphysical emergentist must balance these two opposing consequences of metaphysical emergence for and against applying the precautionary principle, and reaching that balance requires multifaceted analyses of the precautionary principle, a deep understanding of metaphysical emergence, and detailed case studies.

In the next subsection, I expand on the first, pro-caution consequences of metaphysical emergence in the context of biological engineering. The discussion draws on the ideas presented in Chapter 2, and will be around the unpredictability of biological emergent systems. However, I emphasize that this is not the whole story, and the opposing *anti-caution* consequences should also be considered. Ideas presented in Chapter 1 can form the basis of this opposing side. Here is a quick sketch of the idea. In Chapter 1, I showed that even though one cannot predict the outcome and the behaviour of emergent systems with complete accuracy and precision, the emergent patterns might have regularities that allow a coarse-grained general prediction of how that system behaves and what comes out of it. A coarse-grained prediction might be enough to show that the risks of applying a new technology are low and thus, form the basis of an argument against applying

the precautionary principle, even if a fully precise and detailed prediction of what comes out of that technology is not possible. I do not further develop this argument here, and instead, focus on the other, pro-caution argument.

5.2.1.1 Emergence compromises risk assessment and risk control in biological engineering

Biological engineering provides many cases for applying the precautionary principle. The field is associated with catastrophic environmental and health hazards (Asante, 2008; Wintle et al., 2017), and we have strong reasons to believe that the list of these hazards can be much longer, and the probabilities assigned to each hazard can be significantly different from our assessments. It is no surprise, therefore, to see the precautionary principle suggested or applied time and again in many cases of biological engineering. Several academics argue for applying various forms of the precautionary principle in biological engineering (Asante, 2008; Aslaksen et al., 2006; Baum & Wilson, 2013; Craig et al., 2008; Myhr, 2010), and the principle has guided the regulations that ban cultivation or import of genetically modified crops in many countries all around the world. In this sub-section, I expand on the reasons for applying the precautionary principle in biological engineering in light of the discussions presented in Chapter 2.

In Chapter 2 I argued that because of the emergent nature of biological systems, a biological engineer has to take recourse to irrational methods. We saw there that one major reason for taking recourse to irrational methods was that the developer could not predict the outcomes of her rational designs. This has dire consequences for risk assessment and risk control. A developer who cannot predict how her system behaves also cannot anticipate how her system could misbehave. And a developer who relies on chance-based irrational methods to set things right, also has to rely on chance if things go wrong. Both her risk assessment and risk control are fundamentally compromised. Of course, engineers have some limited mechanistic insight that allows them to assess some risky aspects of their technologies accordingly (Wintle et al., 2017). However, even for the known threats, they do not know exactly how probable and serious they are. Complexity of biological systems and the irrational nature of biological engineering makes it very hard to assess the scale and the probability of these threats.

In theory, in some cases one can use experiments as opposed to mechanistic knowledge to identify risks by trying and examining many different scenarios. Seat belt crash tests are examples of such approach to risk assessment. But risk assessment by experimenting has low chance of success, partly because one's random sampling might not be large enough to contain all the risky scenarios, and partly because one does not know exactly what to look for when searching for risks.

Contrary to the functional behaviours, the problematic behaviours are multifarious and are not known a priori and therefore, are much harder to catch. Some critically risky scenarios might never be encountered in trials, and some might appear but completely escape the developers' attention. Scientists have practically experienced this in drug development, another field of interaction with biological systems. While drug developers guess some of the side effects based on the drug's mechanism of action and observe some others during clinical trials, there are still many side effects that do not show themselves until the drug is widely used by the general public.

Besides risk identification, risk control is also impeded by the irrational model of development in biological engineering. Even if the developers somehow come up with the full list of risky scenarios, their ability to prevent or remedy those scenarios would be limited. Lacking a mechanistic understanding of their system, they do not know a priori what modifications they should make in the system to account for the risks. They should spin the wheel of fortune by trial and error, and have their fingers crossed. And if they are lucky enough to find a modification that does the job, they have only probabilistic, and not mechanistic evidence for its reliability. It has worked in our limited test cases, they must say, so we hope that it will work every time, everywhere.

In Chapter 2, I examined examples of genetic circuits designed by Gardner et al., and Elowitz and Leiber. There was considerable uncertainty and ambiguity even in those simple well-understood genetic systems. Today CRISPR technology allows us to edit genes in the context of three billion base pairs of the human genome. Just imagine the uncertainty and the ambiguity. In 2014, Maddalo et al. equipped a virus with the CRISPR components, and infected mice with the virus via inhalation to induce them with lung cancer. The aim was to make a model for human cancer research (Maddalo et al., 2014). Can the virus infect the human lung as well? Andrea Ventura, a lead author of the work, responds "[t]he guides [RNA sequences that take the technology to its target genes] are not designed to cut the human genome", he continues, "but you never know" (Ledford, 2015:21).

In biological engineering we "never know" and that is the problem. We do not know exactly how the system works, we do not know what can possibly go wrong, and we do not know what we should do about it. In engineering language, the irrational model of biological engineering compromises the *safe design* principles. Looking at various discipline-specific guidelines, the philosopher Hansson (2006) summarises four major safe design principles that apply to any engineering development. The first principle is *inherently safe design*. This principle dictates that the designer should make a system that is inherently safe so that implementing additional safety measures become unnecessary. For example, a machine should ideally be built out of

noninflammable material so that no fire-extinguishing mechanism is needed. It is impossible to apply this principle to biologically engineered systems because one can eliminate the source of a risk only if one is aware of the risk and knows its sources. But as discussed above, in biological engineering we often do not know many of the potential risks and their underlying causes. Therefore, we lack the necessary insight to apply this principle, and the inherent risks remain there in the biologically engineered systems.

The second principle is implementing *safety factors*. The principle tells that the engineer should always allow for unusual extra pressure on the system. For example, a bridge should be able to tolerate much more weight than its estimated usual load. Considering the unpredictability of biological systems, applying this principle can be very helpful, but it is hard to find actual cases of its application. Some hypothetical cases could be ensuring that a viral vector, a genetic circuit, or any genetic manipulation remains safe even if the mutation rate gets much higher than what is expected. Undoubtedly, some requirement along this line can contribute to the safety of genetic engineering. However, the implemented safety factors would themselves be unpredictable due to their biological nature, and might behave differently from one system to another, and from one environment to the next. Such safety measures cannot be fully reliable.

The third principle advises implementing *negative feedback* mechanisms that shut down the system under particular conditions, for example, in case of system failure or after some expiration period. Ordinary electric fuses are good simple examples of such mechanisms. This principle has attracted attention in biological engineering, and several negative feedback mechanisms are suggested for different systems and scenarios. For example, synthetic biology genetic circuits are designed to switch off CRISPR technology if need be (Pineda et al., 2019). Pineda et al. (2019) suggest that using these circuits are indeed the best path towards developing safe CRISPR-based therapeutics.

Undeniably, biological negative feedback safety switches have some promise as safety measures. But these switches are themselves instances of biological engineering and therefore, come with all the uncertainties of such systems. The behaviours of synthetic biology circuits are fluctuating even in controlled simplified experimental setups. How can we be sure about their reliable performance in unforeseen risk-associated extreme scenarios? Because of this inherent uncertainty, the safety of engineered systems cannot be entirely entrusted to these switches.

The fourth principle is having *multiple independent safety barriers*. The principle says that the engineer should implement several orthogonal safety barriers so that a common cause cannot result in failure of all of them at the same time, and the failure of one barrier does not cascade to

failure of others. This principle is recognised in biological engineering, particularly for containment of gene drives (Akbari et al., 2015; Heitman et al., 2016). For instance, a group of prominent gene drive researchers have suggested four independent types of confinement measures to prevent unwanted release and propagation of gene drives in nature. The first type is *molecular* (or *biological*), and it refers to biological designs to confine gene drive activation and propagation strictly to the situations envisioned by the experimenter. The second type is *reproductive,* and it refers to manipulating the gene drive carrier species so that they cannot reproduce with the wild type strains. The third type is *ecological*, and it refers to running the ecological experiments with gene drive carrier species outside of the habitat of the natural corresponding species. And the fourth type is *physical*, and it refers to using multiple physical barriers to prevent unintentional release of the carrier organism to nature (Akbari et al., 2015).

However, after listing all these orthogonal barriers, these researchers suggest that usually two barriers should be enough and under certain circumstances, even one could suffice. Therefore, the principle of multiple barriers is advertised, but not fully enforced. One reason could be that it is not always possible to have all the different types of barriers in place. For example, consider the proposal of using gene drives to exterminate malaria-transmitting mosquitoes. The goal of the project is to eliminate or manipulate the natural population of such mosquitoes via propagating specific genes in their population by releasing gene carrying mosquitoes into their habitat. For such a project the ecological, the reproductive, and the physical types of confinement measures are inherently off the table. We are left only with the biological measures.

When we cannot have orthogonal barriers of different biological and non-biological types, can we have multiple orthogonal biological safety measures? From our discussion of indecomposability of biological systems in Chapter 2, it follows that independent biological safety measures are most probably not possible. As indecomposable systems, all parts of the biological system are inter-connected and there are no two parts of the system that are completely independent. So, a cause may affect the state of the system in a way that disables multiple biological mechanisms. Or the failure of one of the safety mechanisms may change the state of the system in a way that results in the collapse of the others as well.

The conclusion is that because of the emergent nature of biological systems, biological engineers' ability for risk assessment and risk control of biological products is severely limited. Moreover, the conclusions of Chapter 2 tell us that we should not be optimistic that engineers will eventually overcome these limitations. As emergent systems, biological systems are not going to lend themselves to the kind of mechanistic understanding and rational development that is needed

for safe implementation of biological engineering. We should be cautious, or so this pro-caution side of the story goes.

5.2.1.2 Naïve precaution is unjustified

It is important to note that although unpredictability of biological systems provides strong reasons in support of the application of the precautionary principle, naïve precaution is not justified. One of the most important reasons is the so-called objection of incoherence against the precautionary principle (Bodansky, 1991). The idea is that the precautionary principle depicts the situation as if we are deciding between risk and precaution. But in reality, we usually decide between two risks. Being precautious and not accepting one of these risks means that we are embracing the other. A telling example of this risk-risk scenarios is the case of genetically modified food in Africa. In 2002, Zambian government rejected tons of food aid from the US because they contained genetically modified kernels. At the very time, the World Health Organization estimated that 35,000 Zambians would die of starvation if more food could not be provided (Sunstein, 2005). One can cite the precautionary principle and suggest rejecting the food aid to avoid the risks of genetically modified food. But rejecting the food aid means embracing the risk of hunger and the precautionary principle also applies to that risk. Therefore, one can cite the very same principle and suggest avoiding the risk of hunger by accepting the food aid. Thus, so the objection goes, the precautionary principle pushes both for and against accepting the food aid and is incoherent.[31]

The objection of incoherence holds in many cases of biological engineering, but it does not hold everywhere. For example, the same engineering techniques used to develop high-yield crops can be used to develop roses with shinier colours and bonsai trees with cuter looks. In such cases, it is only one side of the problem, i.e. the side of applying the technology, that is associated with large environmental risks. In such cases, the precautionary principle applies only to one side of the problem and will not generate contradictory outcomes.

Even in cases where we do have a risk-risk scenario, a supporter of the precautionary principle may eliminate the risk of one side by suggesting alternative non-biological ways to address the problem at hand. In this way, the precautionary principle would apply only to one side of the problem and will not be incoherent. For example, one may suggest that correcting the food distribution and consumption practices can address the hunger problem much more efficiently and safely than using genetically modified crops. The United States is one of the largest suppliers of genetically modified food to Africa. At the same time, according to the United States

[31] Some like Resnik (2021) and Steel (2015) have argued for versions of the precautionary principle that supposedly avoid the objection of incoherence. I leave the analysis of these ideas for another place.

Department of Agriculture, 30 to 40 per cent of the food supply in the US is simply wasted. Only in 2010, the retail and consumer level food waste in the US amounted to 133 billion pounds and 161 billion dollars' worth of food (USDA, 2023). It would be more reasonable, the precautionary principle would suggest, to improve the food consumption and distribution patterns in countries like the US, and redistribute the surplus wasted food to those who need it, rather than exposing the African population and environment to unknown risks of genetically modified crops.

Nonetheless, there can be cases in biological engineering where the precautionary principle applies to both sides of a risk-risk problem and turns out to be incoherent. A telling example is military research on biological engineering. The potential use of biologically engineered tools as a means of terror and weapons of war makes researching these technologies a pragmatic necessity for the defence organisations of all countries. Governments either actively pursue those weapons or want to have a good grip over these technologies to be aware of their military potentials and be able to counter their threats. This tendency is clear in the funding sources for biological engineering projects in the United States. Several biological engineering research programs for developing agricultural tools are not funded by the US Department of Agriculture, but indeed by the Defence Advanced Research Projects Agency (DARPA) that is the agency of the United States Department of Defence responsible for developing technologies for military use (Wintle et al., 2017). A country that does not actively research dual purpose biological technologies puts its nation in the great potential danger of becoming victim of biological violence. A country that does engage in such research, poses its nation and the whole globe to the potential threats of these dangerous technologies. Here, the precautionary principle advises both for and against engaging in this type of research and therefore, becomes incoherent.

The objection of incoherence shows that the precautionary principle could not be applied naively. It is not enough to refer to the emergent nature of biological systems and push for shutting down the whole business of biological engineering. Such decisions need careful and all-rounded case studies. And in debates around these studies, the opposing argument based on coarse-grained predictability of emergent systems that I proposed above could be one of the tools in the repertoire of those who disagree with the precautious.

5.2.2 Beauty as an emergent phenomenon

I propose that we can characterise at least some cases of beauty as metaphysically emergent structural phenomena, and that that characterisation will allow us to attribute epistemic value to aesthetic judgements. In short, I propose that at least in some cases, experiences of beauty are

experiences of emergence. Let me expand a bit on this proposal with focus on scientific aesthetic judgments.

Throughout the history of science, aesthetic values have been one of the important non-empirical criteria for theory choice. A long list of great scientists have relied on beauty as a source of justification and assumed that their theories should be true because they were beautiful (Ivanova, 2017). At first sight, however, there is no relation between the aesthetic value of a theory and its epistemic value. So why should one think a beautiful theory should also be true? This doubt finds strong support in the recent literature that is critical of relying on aesthetic judgements in science (Ellis & Silk, 2014; Hossenfelder, 2018). However, I suggest that understanding beauty as a metaphysically emergent systemic character takes it out of the eyes of the beholder and puts it in the system as an objective fact, thus allowing it to have epistemic value. From this perspective, when someone says that Pergolesi's Stabat Mater is beautiful, she does not merely express her sentiments towards that piece but points to an objective, real, and perhaps causally effective metaphysically emergent character of that musical system. Similarly, when a scientist says that a theory is beautiful, she does not merely state her sentiments towards that theory, but she is pointing to an objective, real, and perhaps causally effective character of the systems described by that theory. As aesthetic judgments point to objective facts, they can have truth value because they may hit or miss in their pointing. Therefore, there can be something "right" about finding Pergolesi's Stabat Mater, or Einstein's theory of relativity, beautiful. And beauty might not be a matter of taste after all.

Establishing that beauty is a metaphysically emergent phenomenon, and showing the epistemic value of aesthetic judgments based on this emergentist characterisation of beauty is a hefty project. Here I sketch only some preliminary remarks. In what follows I first try to briefly substantiate the claim that ordinary and scientific aesthetic judgments point to objective systemic facts. Then I try to quickly respond to the important objection that aesthetic judgments widely disagree and, therefore, are subjective and unreliable.

5.2.2.1 Aesthetic judgments point to objective systemic characters

From the emergentist perspective, beauty is a higher-level emergent character of a holistically harmonious system. One cannot break this holistic property down and attribute it to this or that single element of a beautiful object. There is no single colour or brush stroke that makes a beautiful painting, beautiful. What makes the paining beautiful are the harmonious relations that exist between all the different parts and aspects of the painting. Every aspect is related to every other aspect through convoluted, interdependent relations. The balance that exists between all aspects

of the painting makes the painting beautiful as a whole. Of course, the painting will be deemed beautiful only if an observer appreciates and enjoys those harmonies and so, judgments of beauty partly depend on subjective appreciation. Nonetheless, the holistic harmonies that are deemed beautiful are objective emergent characters of the painting.

If this emergentist characterisation of beauty is correct, we should see that human aesthetic judgments are sensitive to systemic characters of beautiful systems. In other words, seeing beauty should correspond to the beautiful system having some holistic systemic characters. Finding this sensitivity and correspondence is partly a matter of empirical investigation, and in fact, there is already some empirical evidence that shows human aesthetic judgment is sensitive to some systemic characters. One example is the aesthetics of fractals. Fractals are special type of images that are produced by recursive repetition of a particular pattern, and are widely used to model natural patterns (Mandelbrot, 1982). There is empirical evidence that humans detect fractal images and show positive responses to these special type of images (Forsythe et al., 2011; Hagerhall et al., 2008; Joye, 2005, 2006). The positive response is so strong that some have suggested using fractal images as therapeutic tools (Taylor, 2006). Humans not only show preference for fractal images, but even more, they show specific aesthetic preference for some particular sub-types of fractals. An important numerical descriptor of fractals is their fractal dimension or *D*. In general, D is a measure of how details of fractal patterns change at various scales. Forsythe et al (2011) show that D combined with a numerical estimate of image complexity can predict human aesthetic responses to fractal images. Preference for some special sub-type of fractals is not just a matter of public preference, but is also observed in high art. According to an analysis of video footages of Jackson Pollock at work, Pollock was finetuning his painting to have a certain D. Further analysis of his artwork showed that the D values of his works increased over a 10-year period (Taylor et al., 1999; Taylor, 2002), and this discovery was then used to de-authenticate some alleged works of Pollock (Taylor et al., 2007).

These studies do not show that human aesthetic judgment is simply a "D-detector". Rather, we should see D values to numerically represent some systemic characters of fractals which are appealing to human aesthetic sense. This interpretation finds support in other studies that show humans aesthetic judgments are sensitive to the balance between complexity and understandability. Berlyne's pioneering studies show that humans desire a balance between complexity and understandability and therefore find systems of a certain level of complexity beautiful (Berlyne, 1970, 1973). Some later studies confirmed this hypothesis (Kaplan, 1995; Kaplan & Kaplan, 1989). On the one hand these studies show a human psychological preference. But on the other hand,

they show that human aesthetic sense is sensitive to the level of the complexity of the system, which is a holistic, objective character of the system.

In the same line, there is evidence that the scientific sense of beauty is also rooted in objective systemic facts. A telling example is chemical beauty. One area where scientists make judgements of beauty is in medicinal chemistry which is the science of designing chemical compounds with pharmaceutical effects. Some in the field believe that experienced medicinal chemists develop a sense of chemical aesthetics that allows them to identify druggable molecules (Bickerton et al., 2012; Leeson, 2012). By looking at a molecule, an experienced medicinal chemist can tell its promise to make a successful drug. The medicinal chemists can see the "attractive" molecules, although they may not be able to put their finger on exactly what makes those molecules attractive. Lipinski, a prominent medicinal chemist, suggests that this is the only unique skill that a medicinal chemist contributes to a drug discovery team (Lipinski, 2009).

The claim that medicinal chemists' sense of beauty is rooted in objective facts about the molecules finds support in studies that compare the results of computational methods of detecting druggable molecules with the chemists' sense of attractiveness. There are several computational methods to assess drug-likeness. One of the most successful algorithms to identify drug-likeness is the one developed by Bickerton et al (2012). In their paper entitled "quantifying the chemical beauty of drugs", Bickerton et al compared the performance of their algorithms not only with other computational methods, but also with experts' intuitions about the molecules. They asked 79 chemists to collectively categorize 17,117 molecules (approximately 200 molecules by each chemist) into "attractive" or "non-attractive" groups just by looking at the structure of the molecules. Then they compared the mean quantitative drug-likeness score obtained from their algorithm between the attractive and non-attractive groups. The attractive molecules showed a significantly higher average drug-likeness score compared to the non-attractive molecules, which shows that there is a good correlation between the chemists' sense of aesthetics and objective quantitative characterisation of molecular structure.

Bickerton et al's quantitative scores are results of a function that combines numerical values of eight different chemical descriptors such as the number of rotatable bonds (as a measure of flexibility), or octanol-water partition co-efficient (as a measure of lipophilicity). The unique way to combine these eight numerical descriptors to generate the score is obtained by training the function against hundreds of drug molecules. In other words, the score combines various facts about a molecular structure into a holistic single number describing the molecule. We do not know how the chemists' intuitions work. But whatever happens behind the scenes in the brains of

medicinal chemists, the outcome is that the chemists' subjective qualitative aesthetic sense correlates with a quantitative holistic description of the molecular structure.

In summary, there is some evidence that ordinary and scientific aesthetic judgments are sensitive to some holistic structural characters of beautiful systems.

5.2.2.2 Disagreements in aesthetic judgments

One important objection against the objectivity of aesthetic judgments goes that the wide disagreements in aesthetic judgments show that these judgments are merely subjective sentiments and cannot be rooted in objective facts. We face such disagreements in our day-to-day life (who chooses the music in the car?), and also see them in professional aesthetic judgments. For example, Takaoka et al (2003) show that medicinal chemists intuitions on drug-likeness (or molecular "attractiveness") show only a modest 0.5 to 0.6 correlation coefficient. If chemists' intuitions are rooted in objective facts, so the objection goes, how is it possible that they vary so significantly? Why cannot scientists convince each other and agree on one aesthetic conclusion?

An emergentist would have a response. She could say that the holistic nature of beauty makes it particularly hard to substantiate aesthetic judgements and convince others. One somehow sees the general harmony that exists at the systemic level without being able to break it down to different facts about the system. The two sides of an aesthetic debate cannot clearly support their positions by pointing to this or that particular fact about the system and explain exactly why they see beauty. But this does not mean that these judgements are simply matters of taste with no objective basis. The disagreements and the difficulties in citing convincing evidence in support of one's judgment are simply side effects of the emergent nature of beauty.

But this leads to another question. If we cannot convince everyone to accept a common aesthetic judgement, whose judgment should we trust? Even if we accept that aesthetic judgments are rooted in objective facts, don't the wide disagreements mean that we cannot be sure which judgment is the objectively correct one? I think there are two ways to go around this problem. First, we should not give all aesthetic judgments the same weight. The emergent nature of beauty and its irreducibility to simple factors makes aesthetic judgements a proprietary area of the experienced. The more experienced, the more reliable the judgment. In fact, the ability to make reliable aesthetic judgments is one of the important capabilities that a true master of an art should develop. A chess master, for example, should be able to see a "beautiful" move that wins the game. Not all chess players are equally qualified to see such beautiful moves. Seeing those moves is the proprietary ability of true masters. The same holds for scientific aesthetic judgments. For example, in our chemical case, we should note that not all medicinal chemists are equal. Chemists vary

greatly in the amount of their experience, their natural talents, and their training. It is unsurprising, therefore, to see great variability in the aesthetic judgments of a large cohort of chemists. To find the reliable aesthetic judgment, we should rank the chemists according to their level of mastery and give weight to their judgments accordingly.

Second, we can refer to the collective sense of beaty. Reliable judgments of beauty in science are those that are not merely the opinion of some individual scientists, but are the collective general judgement of the field despite all differences (Butterfield, 2019). Going back to the case of medicinal chemists, we saw above that there is good evidence that this collective aesthetic intuition correctly identifies druggable molecules, despite all the differences between induvial judgments. And finally, we can combine the first and the second solutions together and come with weighted collective aesthetic conclusions. So, in short, the wide variation in aesthetic judgments does not necessarily mean that we cannot have reliable aesthetic conclusions.

In summary, there is some evidence that aesthetic judgments are sensitive to systemic emergent facts and are thus not merely subjective. This objectivity allows these judgments to have real epistemic value. Also, from the practical perspective, there are ways to solve the problem of disagreements in aesthetic judgments and come to reliable aesthetic conclusions. Therefore, aesthetic judgments can potentially be an important part of our epistemic toolbox. Of course, this is only a preliminary remark that needs much further development and substantiation.

I want to conclude my dissertation with beauty. So, let me stop philosophising here, and close this desertion with the first and the last stanzas of *Pythoness* by Kathleen Raine (1908-2003), a Cambridge alumnus, that, to my reading, beautifully describes metaphysical emergence. Even if my future work fails to establish that beauty is emergent, these stanzas are enough to establish that emergence can be beautiful:

I am that serpent-haunted cave
Whose navel breeds the fates of men.
All wisdom issues from a hole in the earth;
The gods form in my darkness, and dissolve again.
…
I am that feared and longed-for burning place
Where man and phoenix are consumed away,
And from my low polluted bed arise
New sons, new suns, new skies.

References

Aaronson, S. (2002). Book Review on A New Kind of Science. *Quantum Information and Computing*, *2*(5), 410–423.

Ahteensuu, M. (2007). Defending the precautionary principle against three criticisms. *Trames*, *11*(4), 366–381.

Aizawa, K. (2013). Multiple realization by compensatory differences. *European Journal for Philosophy of Science*, *3*(1), 69–86.

Akbari, O. S., Bellen, H. J., Bier, E., Bullock, S. L., Burt, A., Church, G. M., Cook, K. R., Duchek, P., Edwards, O. R., Esvelt, K. M., & others. (2015). Safeguarding gene drive experiments in the laboratory. *Science*, *349*(6251), 927–929.

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.

Anderson, P. W. (1972). More is different. *Science*, *177*(4047), 393–396.

Asante, D. K. A. (2008). Genetically modified food-The dilemma of Africa. *African Journal of Biotechnology*, *7*(9).

Aslaksen, I., Natvig, B., & Nordal, I. (2006). Environmental risk and the precautionary principle:"late lessons from early warnings" applied to genetically modified plants. *Journal of Risk Research*, *9*(03), 205–224.

Austin, P. (2021). Theory of language: a taxonomy. *SN Social Sciences*, *1*(3), 1–24.

Auyang, S. Y. (1998). *Foundations of complex-system theories: in economics, evolutionary biology, and statistical physics*. Cambridge University Press.

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

Balari, S., Antonio, B.-B., Longa, V. M., & Lorenzo, G. (2013). The fossils of language: What are they? Who has them? How did they evolve? In C. Boeckx & K. K. Grohmann (Eds.), *The Cambridge handbook of biolinguistics* (pp. 489–523).

Balari, S., Lorenzo, G., & González, G. L. (2013). *Computational phenotypes: towards an evolutionary developmental biolinguistics* (Vol. 3). Oxford University Press.

Barwich, A.-S. (2021). Imaging the living brain: An argument for ruthless reductionism from olfactory neurobiology. *Journal of Theoretical Biology*, *512*, 110560.

Bates, E., Bretherton, I., & Snyder, L. S. (1988). *From First Words to Grammar: Individual Differences and Dissociable Mechanisms*. Cambridge University Press.

Bates, E., Elman, J., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1998). Innateness and emergentism. *A Companion to Cognitive Science*, 590–601.

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.

Baum, S. D., & Wilson, G. S. (2013). The ethics of global catastrophic risk from dual-use bioengineering. *Ethics in Biology, Engineering and Medicine: An International Journal*, *4*(1).

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. (2002). Downward causation and the autonomy of weak emergence. *Principia*, *6*(1), 5.

Bedau, M. A. (2008). Is weak emergence just in the mind? *Minds and Machines*, *18*(4), 443–459.

Benítez-Burraco, A. (2020). The golden mean: A systems biology approach to developmental language disorders. *Pragmalinguistica*, *2*, 30–44.

Benítez-Burraco, A., & Nikolsky, A. (2023). The (co) evolution of language and music under human self-domestication. *Human Nature*, *34*, 229–275.

Berlekamp, E. R., Conway, J. H., & Guy, R. K. (1982). *Winning Ways for Your Mathematical Plays Vol 2*. Academic Press.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Attention, Perception, and Psychophysics*, *8*, 279--286.

Berlyne, D. E. (1973). *Aesthetics and psychobiology*.

Berto, F., & Tagliabue, J. (2017). Cellular Automata. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017). Metaphysics Research Lab, Stanford University.

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. MIT press.

Berwick, R. C., & Chomsky, N. (2019). All or nothing: No half-Merge and the evolution of syntax. *PLoS Biology*, *17*(11), e3000539.

Bickerton, G. R., Paolini, G. V, Besnard, J., Muresan, S., & Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. *Nature Chemistry*, *4*(2), 90–98.

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*.

Bitbol, M. (2007). Ontology, matter and emergence. *Phenomenology and the Cognitive Sciences*, *6*(3), 293–307.

Bodansky, D. (1991). Scientific uncertainty and the precautionary principle. *Environment: Science and Policy for Sustainable Development*, *33*(7), 4–44.

Bognar, G. (2011). Can the maximin principle serve as a basis for climate change policy? *The Monist*, *94*(3), 329–348.

Breuker, C. J., Debat, V., & Klingenberg, C. P. (2006). Functional evo-devo. *Trends in Ecology & Evolution*, *21*(9), 488–492.

Brier, G. (1950). Verificationn of forcasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3.

Buss, S., Papadimitriou, C. H., & Tsitsiklis, N. J. (1990). On the predictability of coupled automata: An allegory about Chaos. *Proceedings [1990] 31st Annual Symposium on Foundations of Computer Science*, 788–793.

Butterfield, J. (2019). Lost in Math? A review of'Lost in Math: How Beauty Leads Physics Astray', by Sabine Hossenfelder. *ArXiv Preprint ArXiv:1902.03480*.

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.

Castellani, E. (2002). Reductionism, emergence, and effective field theories. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *33*(2), 251–267.

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. 245–254). Oxford University Press.

Charbonneau, P. (2017). *Natural Complexity A Modelling Handbook*. Princeton University Press.

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.

Chen, Y., Li, Y., Narayan, R., Subramanian, A., & Xie, X. (2016). Gene expression inference with deep learning. *Bioinformatics*, *32*(12), 1832–1839.

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M. M., & others. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, *15*(141), 20170387.

Chomsky, N. (1977). *Essays on form and interpretation*. Elsevier Noth Holland.

Chomsky, N. (2001). *Beyond explanatory adequacy*. MIT press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, *36*(1), 1–22.

Cordovil, J. L., Santos, G. C., & Symons, J. (2022). Reconciling Ontic Structural Realism and Ontological Emergence. *Foundations of Science*, 1–20.

Craig, W., Tepfer, M., Degrassi, G., & Ripandelli, D. (2008). An overview of general features of risk assessments of genetically modified crops. *Euphytica*, *164*(3), 853–880.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach*. Longman.

Darley, V. (1994). Emergent phenomena and complexity. *Artificial Life*, *4*, 411–416.

Dasgupta, S. (2011). The bare necessities. *Philosophical Perspectives*, *25*, 115–160.

Davies, P. C. W. (2004). Emergent biological principles and the computational properties of the universe: Explaining it or explaining it away. *Complexity*, *10*(2), 11–15.

Dawkins, R. (2016). *The selfish gene*. Oxford University Press.

Deacon, T. W. (2003). The Hierarchic Logic of Emergence: Untangling the Interdependence of Evolution and Self-organization. In B. H. Weber & D. J. Depew (Eds.), *Evolution and learning: The Baldwin effect reconsidered* (p. 273). MIT Press.

Deacon, T. W. (2005). Language as an emergent function: some radical neurological and evolutionary implications. *Theoria. Revista de Teor'ia, Historia y Fundamentos de La Ciencia*, *20*(3), 269–286.

Deacon, T. W. (2006). Emergence: The hole at the wheel's hub. In P. Clayton & P. C. W. Davies (Eds.), *The re-emergence of emergence: The emergentist hypothesis from science to religion* (pp. 111–150). Oxford University Press.

Deacon, T. W. (2007). Three levels of emergent phenomena. In N. Murphy & W. Stoeger (Eds.), *Evolution and emergence: Systems, organisms, persons* (pp. 88–110). Oxford University Press.

Deacon, T. W. (2014). The Emergent Process of Thinking as Reflected in Language Processing. In D. Schoeller & V. Saller (Eds.), *Thinking thinking Practical radical reflection* (pp. 136–159). Verlag Karl ALber.

Deacon, T. W., & Cashman, T. (2012). Eliminativism, complexity, and emergence. In *The Routledge companion to religion and science* (pp. 193–205). Routledge.

Degiacomi, M. T. (2019). Coupling molecular dynamics and deep learning to mine protein conformational space. *Structure*, *27*(6), 1034–1040.

Dennett, D. (2008). Real Patterns. In M. Bedau & P. Humphreys (Eds.), *Emergence: Contemporary Readings In Philosophy And Science* (pp. 189–205). MIT Press.

Descartes, R. (1998). *Discourse on method*. Hackett Publishing.

Dewar, N. (2019). Algebraic structuralism. *Philosophical Studies*, *176*(7), 1831–1854.

Donahoe, J. W. (2010). Man as Machine: A Review of "Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience" By C. R. Gallistel And A. P. King. *Behavior and Philosophy*, *38*, 83–101. http://www.jstor.org/stable/41806290

Driesch, H. (1905). *The History & Theory of Vitalism* (C. K. O. (trans.) (ed.)). Macmillan and Co., Limited.

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Harvard University Press.

Dupré, J. (1996). Metaphysical Disorder and Scientific Disunity. In P. Galison & D. Stump (Eds.), *The Disunity of Science* (pp. 101–117). Stanford University Press.

Ellis, G., & Silk, J. (2014). Scientific method: Defend the integrity of physics. *Nature News*, *516*(7531), 321.

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, *403*, 335–338.

Ereshefsky, M. (1995). John Dupré, The Disorder of Things: Metaphysical Foundations of the Disunity of Science. Cambridge, MA and London, England: Harvard University Press 1993. Pp. xii+ 308. *Canadian Journal of Philosophy*, *25*(1), 143–158.

Esfeld, M. (2004). Quantum entanglement and a metaphysics of relations. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *35*(4), 601–617.

Fodor, J. (1974). Special sciences, or the disunity of science as a working hypothesis. *Synthese*, 97–115.

Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. MIT press.

Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Philosophical Perspectives*, *11*, 149–163.

Forrest, P. (2020). The Identity of Indiscernibles. In E. N. Zalta (Ed.), *The {Stanford} Encyclopedia of Philosophy* ({W}inter 2). Metaphysics Research Lab, Stanford University.

Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J., & Sawey, M. (2011). Predicting beauty: fractal dimension and visual complexity in art. *British Journal of Psychology*, *102*(1), 49–70.

Frank, R. M. (2008). The language-organism-species analogy: A complex adaptive systems approach to shifting perspectives on "language." In D. Geeraerts, R. Driven, & J. R. Taylor (Eds.), *Body, language and mind Volume 2: Sociocultural Situatedness* (pp. 215–262). Mouton de Gruyter.

French, S. (2014). Shifting to Structures in Biology and Beyond. In *The Structure of the World: Metaphysics and Representation*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199684847.003.0012

French, S., & Ladyman, J. (2010). In defence of ontic structural realism. In *Scientific structuralism* (pp. 25–42). Springer.

French, S., & Redhead, M. (1988). Quantum physics and the identity of indiscernibles. *The British Journal for the Philosophy of Science*, *39*(2), 233–246.

Frigg, R., & Votsis, I. (2011). Everything you always wanted to know about structural realism but were afraid to ask. *European Journal for Philosophy of Science*, *1*(2), 227–276.

Fujita, H., & Fujita, K. (2022). Human language evolution: a view from theoretical linguistics on how syntax and the lexicon first came into being. *Primates*, *63*(5), 403–415.

Gallistel, C. R., & King, A. P. (2010). *Memory and the computational brain: Why cognitive science will transform neuroscience*. John Wiley & Sons, Ltd.

Ganeri, J. (2011). Emergentisms, ancient and modern. *Mind*, *120*(479), 671–703.

Gardiner, S. M. (2006). A core precautionary principle. *Journal of Political Philosophy*, *14*(1), 33–60.

Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature*, *403*, 339–342.

Gibb, S. (2019). The causal closure principle. In S. Gibb, R. Findlay Hendry, & T. Lancaster (Eds.), *The Routledge handbook of emergence* (pp. 111–120). Routledge.

Gillett, C. (2019). Emergence, downward causation and its alternatives: critically surveying a foundational issue. In S. Gibb, R. Findlay Hendry, & T. Lancaster (Eds.), *The Routledge*

Handbook of Emergence (pp. 99–110). Routledge.

Godfrey-Smith, P. (2003). Scientific Realism. In *Theory and Reality* (pp. 173–189). University of Chicago Press.

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

Griffiths, P. E. (2007). Evo-Devo meets the mind: towards a developmental evolutionary psychology. In *Integrating Evolution and Development: Form Theory to Practice* (pp. 195–225). MIT Press.

Habib, M., & Stacho, J. (2013). Unique perfect phylogeny is intractable. *Theor. Comput. Sci.*, *476*, 47–66.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press.

Hagerhall, C. M., Laike, T., Taylor, R. P., Küller, M., Küller, R., & Martin, T. P. (2008). Investigations of human EEG response to viewing fractal patterns. *Perception*, *37*(10), 1488–1494.

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

Hansson, S. O. (2006). Safe design. *Techné: Research in Philosophy and Technology*, *10*(1), 45–52.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579.

Heitman, E., Sawyer, K., & Collins, J. P. (2016). Gene Drives on the Horizon: Issues for Biosafety. *Applied Biosafety*, *21*(4), 173–176. https://doi.org/10.1177/1535676016672631

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 Amrican Medical Association*, *300*(13), 1580–1581.

Holland, J. H. (1998). *Emergence: From chaos to order*. Oxford University Press.

Horgan, T. (1989). Mental quausation. *Philosophical Perspectives*, *3*, 47–76.

Horgan, T. (1997). Kim on mental causation and causal exclusion. *Philosophical Perspectives*, *11*, 165–184.

Hossenfelder, S. (2018). *Lost in math: how beauty leads physics astray*. Hachette UK.

Hovda, P. (2008). Quantifying weak emergence. *Minds and Machines*, *18*(4), 461–473.

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). *Emergence: A philosophical account*. Oxford University Press.

Humphreys, P. (2016c). Inferential Emergence. In *Emergence: A Philosophical Account*. Oxford University Press.

Humphreys, P. (2016d). 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.

Huneman, P. (2012). Determinism, predictability and open-ended evolution: lessons from computational emergence. *Synthese*, *185*(2), 195–214.

Ilachinski, A. (2001). *Cellular automata: a discrete universe*. World Scientific Publishing Company.

Israeli, N., & Goldenfeld, N. (2006). Coarse-graining of cellular automata, emergence, and the predictability of complex systems. *Physical Review E*, *73*(2), 26203.

Ivanova, M. (2017). Aesthetic values in science. *Philosophy Compass*, *12*(10), e12433.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly*, *32*(127), 127–136.

Jacobs, J., & Seligsohn, M. (1906). Solomon, Seal of. In C. Adler, L. Hühner, & D. Salzberg (Eds.), *The Jewish Encyclopedia* (Vol. 11, p. 448).

Jones, W., Alasoo, K., Fishman, D., & Parts, L. (2017). Computational biology: deep learning. *Emerging Topics in Life Sciences*, *1*(3), 257–274.

Joos, E. (2006). The emergence of classicality from quantum theory. In P. Clayton & P. Davies (Eds.), *The re-emergence of emergence: the emergentist hypothesis from science to religion* (Issue 159, pp. 53–78). Oxford University Press.

Jordan, G. (2003). *Theory Construction in Second Language Acquisition*. John Benjamins Publishing Company. http://ebookcentral.proquest.com/lib/cam/detail.action?docID=623219

Joye, Y. (2005). Evolutionary and cognitive motivations for fractal art in art and design education. *International Journal of Art & Design Education*, *24*(2), 175–185.

Joye, Y. (2006). Some reflections on the relevance of fractals for art therapy. *The Arts in Psychotherapy*, *33*(2), 143–147.

Kaiser, M. I. (2017). The Limits of Reductionism in the Life Sciences. *History and Philosophy of Life Sciences*, *33*(4), 453–476.

Kallestrup, J. (2006). The causal exclusion argument. *Philosophical Studies*, *131*(2), 459–485.

Kandel, E. R., & Hawkins, R. D. (1992). The biological basis of learning and individuality. *Scientific American*, *267*(3), 78–87.

Kantorovich, A. (2003). The priority of internal symmetries in particle physics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *34*(4), 651–675.

Kaplan, S. (1995). Review of the biophilia hypothesis. *Environment and Behavior*, *27*, 801–804.

Kaplan, S., & Kaplan, R. (1989). The visual environment: Public participation in design and planning. *Journal of Social Issues*, *45*(1), 59–86.

Kappraff, J. (2001). *Connections: The geometric bridge between art and science*. World Scientific.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmental disorders. *Developmental Psychology*, *45*(1), 56.

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press.

Kauffman, S. A. (1995). *At home in the universe: The search for laws of self-organization and complexity*. Oxford University Press.

Kauffman, S. A. (2000). *Investigations*. 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. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research*, *52*(1), 1–26.

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.

Kincaid, H. (2008). Structural realism and the social sciences. *Philosophy of Science*, *75*(5), 720–731.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Kitano, H. (2002). Computational systems biology. *Nature*, *420*(6912), 206–210.

Kronz, F. M., & Tiehen, J. T. (2002). Emergence and quantum mechanics. *Philosophy of Science*, *69*(2), 324–347.

Kwok, R. (2010). Five hard truths for synthetic biology: can engineering approaches tame the complexity of living systems? *Nature*, *463*(7279), 288–291.

Ladyman, J. (1998). What is structural realism? *Studies in History and Philosophy of Science Part A*, *29*(3), 409–424.

Ladyman, J., & Ross, D. (2007). *Every thing must go: Metaphysics naturalized*. Oxford University Press on Demand.

Latourelle, J. C., Beste, M. T., Hadzi, T. C., Miller, R. E., Oppenheim, J. N., Valko, M. P., Wuest, D. M., Church, B. W., Khalil, I. G., Hayete, B., & others. (2017). Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. *The Lancet Neurology*, *16*(11), 908–916.

Lecours, A. R., Lhermitte, F., & Bryans, B. (1983). *Aphasiology*. Bailliere Tindall Limited.

Ledford, H. (2015). CRISPR, the disruptor. *Nature News*, *522*(7554), 20.

Leeson, P. (2012). Chemical beauty contest. *Nature*, *481*(7382), 455–456.

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

Lewis, R., & Carroll, F. (2016). Creating seating plans: a practical application. *Journal of the Operational Research Society*, *67*(11), 1353–1362.

Lieberman, P. (2000). *Human Language and Our Reptilian Brain: The Subcortical Bases of Speech, Syntax, and Thought*. Harvard University Press.

Lieberman, P. (2016). The evolution of language and thought. *J. Anthropol. Sci*, *94*, 127–146.

Lipinski, C. A. (2009). Overview of hit to lead: The medicinal chemist's role from HTS retest to lead optimization hand off. In *Lead-Seeking Approaches* (pp. 1–24). Springer.

Litt, A., Eliasmith, C., Kroon, F. W., Weinstein, S., & Thagard, P. (2006). Is the brain a quantum computer? *Cognitive Science*, *30*(3), 593–603.

Liu, Y., Gao, C., Wang, P., Friederici, A. D., Zaccarella, E., & Chen, L. (2023). Exploring the neurobiology of Merge at a basic level: insights from a novel artificial grammar paradigm. *Frontiers in Psychology*, *14*, 1151518.

Lorenz, E. (1972). Does the flap of a butterfly's wings in Brazil set off a tornado in Texas. *Transcript of a Lecture given to the 139th Meeting of the American Association for the Advancement of Science, in Washington, DC*.

Lorenz, E. (1993). *The Essence of Chaos*. University of Washington Press.

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

Lupyan, G., & Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences*, *20*(9), 649–660.

Luu, J., & Palczewski, K. (2018). Human aging and disease: lessons from age-related macular degeneration. *Proceedings of the National Academy of Sciences*, *115*(12), 2866–2872.

MacWhinney, B. (2002). *Language emergence*.

MacWhinney, B. (2005). The emergence of grammar from perspective. *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics*, *95*.

MacWhinney, B. (2009). The emergence of linguistic complexity. In T. Givón & M. Shibatani (Eds.), *Syntactic Complexity* (pp. 405–432). John Benjamins.

MacWhinney, B. (2015). Introduction. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence*. Wiley.

MacWhinney, B., & Bates, E. (1989). *The Crosslinguistic Study of Sentence Processing*. Cambridge University Press.

Maddalo, D., Manchado, E., Concepcion, C. P., Bonetti, C., Vidigal, J. A., Han, Y.-C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E., & others. (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature*, *516*(7531), 423–427.

Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Molecular Pharmaceutics*, *13*(5), 1445–1454.

Mandelbrot, B. B. (1982). *The fractal geometry of nature*. WH freeman New York.

Marcus, G. F. (2008). *Kluge: the haphazard construction of the human mind*. Houghton Mifflin.

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.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (pp. 3–36). Erlbaum.

McKenzie, K. (2017). Ontic structural realism. *Philosophy Compass*, *12*(4), e12399.

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.

Morgan, C. L. (1925). *Emergent evolution*. Williams and Norgate.

Myhr, A. I. (2010). A precautionary approach to genetically modified organisms: challenges and implications for policy and science. *Journal of Agricultural and Environmental Ethics*, *23*(6), 501–525.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450.

Nagel, T. (2012). *Mind and cosmos: why the materialist neo-Darwinian conception of nature is almost certainly false*. Oxford University Press.

Newman, S. A., & Comper, W. D. (1990). 'Generic'physical mechanisms of morphogenesis and pattern formation. *Development*, *110*(1), 1–18.

Newman, S. A., Forgacs, G., & Muller, G. B. (2003). Before programs: the physical origination of multicellular forms. *International Journal of Developmental Biology*, *50*(2–3), 289–299.

Noble, D. (2008). *The music of life: biology beyond genes*. Oxford University Press.

O'Connor, T., & Wong, H. Y. (2005). The metaphysics of emergence. *Noûs*, *39*(4), 658–678.

O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Routledge.

O'Grady, W. (2008). The emergentist program. *Lingua*, *118*(4), 447–464.

O'Grady, W. (2021). *Natural Syntax: An Emergentist Primer V2.2*.

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. *Annu. Rev. Biochem.*, *74*, 867–900.

Oyama, S. (2000). *The ontogeny of information: Developmental systems and evolution* (2th ed., r). Duke university press.

Oyama, S., Gray, R. D., & Griffiths, P. E. (2001). *Cycles of contingency: Developmental systems and evolution*. Mit Press.

Palmer-Brown, D., Tepper, J. A., & Powell, H. M. (2002). Connectionist natural language parsing. *Trends in Cognitive Sciences*, *6*(10), 437–442.

Penrose, R., & Gardner, M. (1989). *The Emperor's New Mind: Concerning Computers, Minds, and The Laws of Physics*. Oxford University Press. https://doi.org/10.1093/oso/9780198519737.001.0001

Peterson, M. (2007). The precautionary principle should not be used as a basis for decision-making. *EMBO Reports*, *8*(4), 305–308.

Piattelli-Palmarini, M. (1990). An ideological battle over modals and quantifiers. *Behavioral and Brain Sciences*, *13*(4), 752–754.

Piattelli-Palmarini, M., & Uriagereka, J. (2004). The immune syntax: the evolution of the language virus. In L. Jenkins (Ed.), *Variation and universals in biolinguistics* (pp. 341–377). Brill.

Pineda, M., Lear, A., Collins, J. P., & Kiani, S. (2019). Safe CRISPR: Challenges and Possible Solutions. *Trends in Biotechnology*, *37*(4), 389–401.

Pinker, S. (1994). *The language instinct: How the mind creates language*. William Morrow.

Plsek, P. E., & Greenhalgh, T. (2001). The challenge of complexity in health care. *BMJ*, *323*, 625–628.

Poeppel, D. (2011). Genetics and language: a neurobiological perspective on the missing link (-ing hypotheses). *Journal of Neurodevelopmental Disorders*, *3*, 381–387.

Poeppel, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology*, *29*(1–2), 34–55.

Poeppel, D., & Embick, D. (2005). Defining the relation between Linguistics and Neuroscience. In *Twenty-First Century Psycho-linguistics: Four Cornerstones* (pp. 103–118). Lawrence Erlbaum.

Poincaré, H. (1952). *Science and hypothesis*. Dover.

Rendell, P. (2011). A universal turing machine in conway's game of life. *2011 International Conference on High Performance Computing & Simulation*, 764–772.

Resnik, D. B. (2003). Is the precautionary principle unscientific? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *34*(2), 329–344.

Resnik, D. B. (2021). *Precautionary reasoning in environmental and public health policy*. Springer.

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

Ritt, N. (2004). *Selfish sounds and linguistic evolution: A Darwinian approach to language change*. Cambridge University Press.

Rivoire, O., Reynolds, K. A., & Ranganathan, R. (2016). *Evolution-Based Functional Decomposition of*

Proteins. 1–26. https://doi.org/10.1371/journal.pcbi.1004817

Ross, D. (2000). Rainforest Realism: A Dennettian Theory of Existence. In D. Ross, A. Brook, & D. L. Thompson (Eds.), *Dennett's Philosophy: A Comprehensive Assessment*. The MIT Press. https://doi.org/10.7551/mitpress/2335.003.0010

Rothschild, L. J. (2006). The role of emergence in biology. In P. Clayton & P. C. W. Davies (Eds.), *The re-emergence of emergence: The emergentist hypothesis from science to religion* (pp. 151–165). Oxford University Press.

Rucker, R. (2003). A New Kind of Science. In *The American Mathematical Monthly* (Vol. 110, Issue 9, pp. 851–861).

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Sandin, P. (2004). *Better safe than sorry:: Applying philosophical methods to the debate on risk and the precautionary principle*. Infrastruktur.

Santos, G. C. (2015). Ontological emergence: How is that possible? Towards a new relational ontology. *Foundations of Science*, *20*(4), 429–446.

Schaffer, J. (2003). Is there a fundamental level? *Noûs*, *37*(3), 498–517.

Shapiro, J. A. (2011). *Evolution: A View from the 21s Century Perspective*. FT Press Science.

Shapiro, S. (2000). Structuralism. In *Thinking about mathematics: The philosophy of mathematics*. OUP Oxford.

Shariati, F., & Khoshnevis, Y. (2020). بررسی تطبیقی نظریه دوگانه‌انگاری نوخاسته (هسکر) و حدوث جسمانی نفس (صدرالمتالهین) در نحوه تعلق یا تکون نفس. *دو فصلنامه تاملات فلسفی*, *10*(24), 289–313.

Shoemaker, S. (2000). Realization and mental causation. *The Proceedings of the Twentieth World Congress of Philosophy*, *9*, 23–33.

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.

Sirois, S., Spratling, M., Thomas, M. S. C., Westermann, G., Mareschal, D., & Johnson, M. H. (2008). Précis of Neuroconstructivism: How the brain constructs cognition. *Behavioral and Brain Sciences*, *31*(3), 321–331.

Smolin, L. (2013). *Time Reborn: From the Crisis in Physics to the Future of the Universe*. Houghton Mifflin Harcourt.

Sober, E. (1999). The multiple realizability argument against reductionism. *Philosophy of Science*, *66*(4), 542–564.

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.

Sperber, D. (2001). In defense of massive modularity. In *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler* (pp. 47–57). MIT Press.

Starr, T. N., & Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Science*, *25*, 1204–1218. https://doi.org/10.1002/pro.2897

Steel, D. (2015). *Philosophy and the precautionary principle*. Cambridge University Press.

Sunstein, C. R. (2005). *Laws of fear: Beyond the precautionary principle*. Cambridge University Press.

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

Tabatabaei Ghomi, H. (2023). Irrational methods suggest indecomposability and emergence. *European Journal for Philosophy of Science*, *13*(1).

Tabatabaei Ghomi, H., & Benítez-Burraco, A. (2023). A philosophical analysis of the emergence of language. *Theoria*. https://doi.org/https://doi.org/10.1111/theo.12507

Takaoka, Y., Endo, Y., Yamanobe, S., Kakinuma, H., Okubo, T., Shimazaki, Y., Ota, T., Sumiya, S., & Yoshikawa, K. (2003). Development of a method for evaluating drug-likeness and ease of synthesis using a data set in which compounds are assigned scores based on chemists' intuition. *Journal of Chemical Information and Computer Sciences*, *43*(4), 1269–1275.

Tang, B., Pan, Z., Yin, K., & Khateeb, A. (2019). Recent advances of deep learning in bioinformatics and computational biology. *Frontiers in Genetics*, *10*, 214.

Taylor, R. P. (2002). Order in Pollock's chaos. *Scientific American*, *287*(6), 116–121.

Taylor, R. P. (2006). Reduction of physiological stress using fractal art and architecture. *Leonardo*, *39*(3), 245–251.

Taylor, R. P., Guzman, R., Martin, T. P., Hall, G. D. R., Micolich, A. P., Jonas, D., Scannell, B. C., Fairbanks, M. S., & Marlow, C. A. (2007). Authenticating Pollock paintings using fractal geometry. *Pattern Recognition Letters*, *28*(6), 695–702.

Taylor, R. P., Micolich, A. P., & Jonas, D. (1999). Fractal analysis of Pollock's drip paintings. *Nature*, *399*(6735), 422.

Tegmark, M. (2000). Why the brain is probably not a quantum computer. *Information Sciences*, *128*(3–4), 155–179.

Tegmark, M., & Wheeler, J. A. (2001). 100 years of quantum mysteries. *Scientific American*, *284*(2), 68–75.

Tetlock, P. E., & Gardner, D. (2016). *Superforecasting: The art and science of prediction*. Random House.

Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *2*(1), 230–265.

USDA. (2023). *Food waste*. https://www.usda.gov/foodwaste/faqs

Van Cleve, J. (1990). Mind--Dust or Magic? Panpsychism Versus Emergence. *Philosophical Perspectives*, *4*, 215–226.

Van Regenmortel, M. H. V. (2004). Emergence in biology. *Proceedings of the Evry Spring School on Modelling and Simulation of Biological Processes in the Context of Genomics*, 123–132.

VanderZwaag, D. (2002). The precautionary principle and marine environmental protection: slippery shores, rough seas, and rising normative tides. *Ocean Development &International Law*, *33*(2), 165–188.

Walsh, D. (2012). Mechanism and purpose: A case for natural teleology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 173–181.

Walsh, D. (2014). The Affordance Landscape: The Spatial Metaphors of Evolution. In G. Barker, E. Desjardins, & T. Pearce (Eds.), *Entangled Life: Organism and Environment in the Biological and Social Sciences* (pp. 213–236). Springer Netherlands. https://doi.org/10.1007/978-94-007-7067-6_11

Walsh, D. (2015a). Object and agent, enacting evolution. In *Organisms, Agency, and Evolution* (pp. 208–229). Cambridge University Press.

Walsh, D. (2015b). *Organisms, agency, and evolution*. Cambridge University Press.

Wang, M., Li, A., Sekiya, M., Beckmann, N. D., Quan, X., Schrode, N., Fernando, M. B., Yu, A., Zhu, L., Cao, J., & others. (2021). Transformative network modeling of multi-omics data reveals detailed circuits, key regulators, and potential therapeutics for Alzheimer's disease. *Neuron*, *109*(2), 257–272.

Warfield, T., & Crisp, T. (2001). Kim's Master Argument. *Nous*, *35*(2), 304–316.

Wayne, A., & Arciszewski, M. (2009). Emergence in physics. *Philosophy Compass*, *4*(5), 846–858.

Weinreich, D. M., Lan, Y., Wylie, C. 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. (1999). How superduper does a physicalist supervenience need to be? *The Philosophical Quarterly*, *49*(194), 33–52.

Wilson, J. M. (2010). Non-reductive physicalism and degrees of freedom. *The British Journal for the Philosophy of Science*, *61*(2), 279–311.

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 Scinces and the Humanities* (pp. 251–306). Brill.

Wilson, J. M. (2021). *Metaphysical emergence*. Oxford University Press.

Wimsatt, W. C. (1994). The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets. *Canadian Journal of Philosophy Supplementary Volume*, *20*, 207–274. https://doi.org/10.1080/00455091.1994.10717400

Wimsatt, W. C. (2000). Emergence as non-aggregativity and the biases of reductionisms. *Foundations of Science*, *5*(3), 269–297.

Wimsatt, W. C. (2006). Aggregate, composed, and evolved systems: Reductionistic heuristics as means to more holistic theories. *Biology and Philosophy*, *21*(5), 667–702.

Wintle, B. C., Boehm, C. R., Rhodes, C., Molloy, J. C., Millett, P., Adam, L., Breitling, R., Carlson, R., Casagrande, R., Dando, M., & others. (2017). Point of View: A transatlantic perspective on 20 emerging issues in biological engineering. *Elife*, *6*, e30247.

Wolfram, S. (1984). Cellular automata as models of complexity. *Nature*, *311*(5985), 419.

Wolfram, S. (2002). *A new kind of science*. Wolfram media.

Wong, H. Y. (2006). Emergents from fusion. *Philosophy of Science*, *73*(3), 345–367.

Worrall, J. (1989). Structural realism: The best of both worlds? *Dialectica*, *43*(1–2), 99–124.

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.

Yanofsky, N. S. (2016). *The outer limits of reason: what science, mathematics, and logic cannot tell us*. MIT Press.

Yoo, J. I., Daugherty, P. S., & O'Malley, M. A. (2020). Bridging non-overlapping reads illuminates high-order epistasis between distal protein sites in a GPCR. *Nature Communications*, *11*(1), 1–12.

Zhu, L., Kim, S.-J., Hara, M., & Aono, M. (2018). Remarkable problem-solving ability of unicellular amoeboid organism and its mechanism. *Royal Society Open Science*, *5*(12), 180396.

Zwirn, H. (2013). Computational irreducibility and computational analogy. *ArXiv Preprint ArXiv:1304.5247*.

Zwirn, H., & Delahaye, J.-P. (2013). Unpredictability and computational irreducibility. In *Irreducibility and Computational Equivalence* (pp. 273–295). Springer.