UNIVERSITÀ DEGLI STUDI DI TORINO

FINO Consortium
Doctoral Thesis

Emergence
Irreducibility, novelty, complexity

Erica Onnis
XXXII cycle

Supervisor: Prof. Alfredo Paternoster

A.A. 2019-2020

Contents

Introduction

In 2008, Paul Humphreys and Mark Bedau edited an original volume titled *Emergence: Contemporary Readings in Philosophy and Science*. In the introduction to the volume, which brings together the analysis of the notion of emergence made by both philosophers and scientists, some examples of alleged emergent phenomena are provided, and those examples range from the properties of certain physical systems to phase transitions, from the phenomenon of life to the mind and consciousness, up to the behaviour of social groups.[1] As highlighted by Humphreys and Bedau, the notion of emergence has a central role in various philosophical and scientific disciplines, and this pervasiveness makes it difficult to produce a univocal definition. Eleven years later, the scenario is clearly similar, except for the further diffusion of the term and the concept of emergence:

> Since the nineteenth century, the notion of emergence has been widely applied in philosophy, particularly in contemporary philosophy of mind, philosophy of science and metaphysics. It has more recently become central to scientists 'understanding of phenomena across physics, chemistry, complexity and systems theory, biology and the social sciences.

This is the statement that introduces, from the back cover, the *Routledge Handbook of Emergence*, published in 2019 by Sophie Gibb, Robin F. Hendry and Tom Lancaster, respectively professors of metaphysics, philosophy of science and condensed-matter

[1] Bedau & Humphreys 2008: 1-2

physics at the University of Durham. The term and the concept of emergence are today more than ever at the centre of the philosophical and scientific debate, and not surprisingly, in recent years, the publications dedicated to this topic have multiplied exponentially. In this introduction I would like to outline a brief contextualization of the debate since, historically, it is possible to recognize two different waves of interest that are rooted in different historical circumstances and theoretical motivations.

The first wave of interest in the concept of emergence features those thinkers that Brian McLaughlin defined "British Emergentists":

> This tradition began in the middle of the nineteenth century and flourished in the first quarter of this century. It began with John Stuart Mill's System of Logic (1843), and traced through Alexander Bain's Logic (1870), George Henry Lewes's Problems of Life and Mind (1875), Samuel Alexander's Space, Time, and Deity (1920), Lloyd Morgan's Emergent Evolution (1923), and C. D. Broad's The Mind and Its Place in Nature (1925).[2]

Despite being in many ways appropriate to unify these thinkers under a single label, the use they make of the concept of emergence is multiple: as pointed out by Joel Walmsley,[3] while Mill and Lewes developed an account of emergence that can be viewed as epistemic for it is correlated to an insufficiency of our knowledge of the natural world, Morgan and Alexander suggest a properly ontological account of emergence, emphasizing the ability of emerging phenomena to exercise novel causal powers. Finally, Broad's work can be considered a middle way between these two conceptions. What British Emergentists had in common, however, was the commitment towards a monist view of the world, namely a metaphysical position implying that the world is not composed by more than just one kind of matter. In particular, British emergentists shared the opinion that reality is composed by physical matter without the addition of any further non-physical elements such as spirits, transcending substances, or other metaphysically controversial entities. Nonetheless, British Emergentists were also aware that admitting physical matter alone would open complex problems about the ontological dignity of apparently non-physical phenomena such as life and mind. For this very reason, they highlighted that despite being exhaustively composed

[2] Mclaughlin 1992: 49.
[3] Walmsley in Onnis 2019.

by and grounded in physical components, those phenomena acquired novel properties in virtue of their particular structure and organisation.

The diffusion of the emergentist theories between the nineteenth and the twentieth century significantly coincides with a historical phase in which physics, chemistry and biology live partially autonomous existences and their unification – however desired – was far to be obtained. It was exactly the possibility of this unification, which became concrete in the 1920s, which represented the main cause of the fall of British Emergentism: according to McLaughlin, the development of quantum mechanics, the explanation of chemical properties through electromagnetism. and the discovery of the molecular structure of DNA opened the way to the general thesis that for any more or less complex natural phenomena a "micro-explanation", namely "the explanation of the behavior of macro-systems in terms of the behavior of their micro-constituents"[4] would be sooner or later available. The alleged availability of micro-physicalistic explanations for each macro-phenomenon coincided, therefore, with the rejection of the emergentist hypothesis. The debate involving British Emergentists was played, therefore, on a purely empirical ground: given some natural phenomena that cannot be explained by physics, it seemed reasonable to hypothesize the existence and causal efficacy of new fundamental and emergent natural forces. However, the scientific discoveries of the first decades of the twentieth century provided good reasons to suppose that the causes of these phenomena could be traced back to more classically physical ones, inflicting a serious blow to the theoretical assumptions of emergentism. It is significant, in this regard, that the latest work clearly attributable to the emergentist movement, Broad's *The Mind and Its Place in Nature*, dates back to 1923,[5] while already in 1922 Niels Bohr was proposing to the scientific community a new and effective atomic model, suggesting how it could be able to explain the chemical properties of the elements of the periodic table.

While in the nineteenth and early twentieth centuries the progress of science had weakened the theory of emergence, it was science that favoured its return and strengthening starting from the seventies of the twentieth century.[6] As attested by the copious scientific debates about emergence, the notion seems useful to describe and understand a series of disparate natural phenomena such the origin of spacetime, quantum entanglement, the

[4] Hüttemann 2004: 24.
[5] Broad's work was published in 1925 but it is composed by the lectures he gave at the Trinity College in 1923.
[6] The paper *More is different*, by the Nobel laureate in physics Philip W. Anderson was originally published in 1972 and was a fundamental publication that would have been mentioned again and again in the debate about emergence and reduction in the following decades.

macroscopic behaviours of molecules and chemical sets, as well as several characteristics of biological and complex systems. Despite the extensive use of the term in all these fields, however, a single definition appropriate to all these contexts is not yet available. As witnessed by Gibb, Hendry, and Lancaster, it appears that philosophers and scientists are using the same term to refer to different things,[7] and this circumstance is relevant and should be taken into consideration to provide a non-prescriptive account of emergence. As pointed out by Mark Bedau,

> […] each conception of emergence must be evaluated independently and on its own merits. It is not competing against other conceptions for the top spot on the podium. Every viable conception would share space on the podium with the other leading candidates for actual and important kinds of emergence.[8]

Keeping in mind this pluralistic position and adopting a descriptive, rather than a prescriptive metaphysical approach, in this work I try to clarify the terms of the debate about emergence providing a description of the phenomenon able to accommodate the different accounts described in the relevant literature. In the first chapter I outline the state-of-the-art of the debate analysing three works recently published and dedicated to the phenomenon of emergence. Those books are Paul Humphreys' *Emergence. A Philosophical Account* (2016), Carl Gillett's *Reduction and Emergence in Science and Philosophy* (2016) and the forthcoming *Metaphysical Emergence* by Jessica Wilson. The literature about emergence is immense, but I decided to start from these three works for two reasons. Firstly, they are three detailed and recent monographies approaching extensively the problem – something which is impossible to do in short papers or book chapters; secondly, the authors explore the literature in their own turn and propose comprehensive taxonomies starting from these examinations and trying to accommodate all the different kinds of emergence present in literature.

Now, reading these three works I notice that despite the differences exhibited by the models of emergence formulated by the authors, at the right level of abstraction it is possible to identify three recurrent criteria that all of them take into account, in one way or another. Emergent phenomena seem to be (i) ontologically irreducible, (ii) epistemologically irreducible, and (iii) novel. However, both irreducibility (ontological as well as

[7] Gibb, Hendry & Lancaster 2019: 2.
[8] Bedau 2013: 92.

epistemological) and novelty are far from being evident and clear concepts. For this reason, I dedicate the second chapter to the relations between emergence and irreducibility, and the third and fourth chapters to those between emergence and novelty. This disproportion – one chapter for irreducibility and two chapters for novelty – depends upon the fact that describing emergence as a non-reducible phenomenon and identifying irreducibility as the mark of emergent entities is a very common strategy providing, however, a *negative* definition which focuses on what emergence is not, rather than on what emergence actually is. Moreover, it may be said – and this is the reading we adopt in this work – that emergent phenomena are irreducible *because* they are in some way novel. Therefore, in our opinion, the third criterion involving novelty is inherently superior to that involving irreducibility. On the one hand, it is a *positive* criterion, identifying substantial features exhibited by emergent phenomena; on the other hand, being novel is the reason why a phenomenon cannot be exhaustively reduced from an ontological point of view. Moreover, very often being irreducible from an ontological point of view is in its own turn the reason why the phenomenon at issue is also epistemologically irreducible. In other terms, the fact that the nature of an emergent phenomenon cannot be deductively predicted (being epistemologically irreducible) frequently descends from the fact that it is ontologically irreducible. And the fact that it is ontologically irreducible depends upon the fact that this phenomenon is novel, in some sense to be defined. There are, therefore, (at least) three levels of analysis of emergent phenomena: their epistemological unpredictability, their ontological irreducibility, and their primary novelty, which is the most promising feature able to clarify the nature of emergents. As the third and fourth chapter elucidate, however, novelty can be – and has been – defined in different ways involving concepts such as nonlinearity, fundamentality, qualitative unprecedented features, and new causal powers.

One way to approach this plurality of definitions could be to identify the most relevant among them, and characterise it – or them – as the necessary and sufficient condition(s) for emergence. In many cases, for instance, this approach favoured the last criterion, namely causal novelty, so that the presence of new causal powers became the necessary condition to have emergence.[9] However, this approach can be short-sighted because these features – nonlinearity, fundamentality, qualitative and causal novelty – all together provide a picture of emergent phenomena that is better than what just one of them could ever do. Moreover,

[9] Jessica Wilson (forthcoming) is a clear example of this stance, but also other scholars can be mentioned. See for instance O'Connor 1994, Bedau 1997, Chalmers 2006 and Gillett 2016.

it is not a coincidence that these qualities have been all related to the concept of novelty: all of them confer to the entity exhibiting them some newness; moreover, they are somehow related among each other and they often imply one another, so they are not mutually exclusive. Another approach which takes into account the importance of all these features, therefore, could be to take all the criteria as necessary and sufficient conditions for emergence. To be emergent, an entity must exhibit nonlinearity, qualitative novelty, fundamentality, and causal efficacy (being at the very same time also epistemologically and ontologically irreducible). However, it is immediately evident that also this strategy cannot be appropriate because it would be excessively strict, excluding certain phenomena exhibiting just some of those features but not all of them. A last remark may be the following. What does it mean to be qualitative novel, fundamental, or causally efficacious is not something that can be defined once and for all. These notions can be assessed in different ways, for they are terms of art whose meanings are related to specific metaphysical frameworks. For this reason, relying in a definitive way upon just some of these notions, as well as rejecting them without challenging them, can be ill-considered. Emergence should not be thought as an unusual natural phenomenon that has to be accommodated in a pre-existent metaphysical view of the world. Its role is more important. Emergence suggests that this metaphysical view should be integrated with new principles and this implies that the terms of the debate has to change to some extent.

In light of this, our suggestion is that the problem of the identification of the relevant criteria for emergence should be approached in the following twofold way. First, it is necessary to evaluate some renovate meanings of novelty in light of emergence – but the same can be said about reduction. As we will see, the nature of certain (emergent) phenomena suggests that the notions of fundamentality and causal efficacy may have been intended in a relatively too strict way in the past. Emergence encourages therefore to widening these notions and, by doing that, to broadening our ontologies. Secondly, as for the problem of the identification of the relevant criteria for emergence, it is not necessary to choose some sufficient and necessary conditions among those mentioned so far, defining them as the absolute mark of emergence. It might be more reasonable to approach this problem in a way which is similar to the approach to natural kinds suggested by Richard Boyd's Homeostatic Property Cluster (HPC) theory. Boyd' theory states that certain natural kinds, given their complexity and the inexactness of the corresponding sciences, cannot be defined by any set of necessary and sufficient conditions but rather by a "'homeostatically' sustained clustering"[10] of properties and relations which constantly but contingently co-occur in nature. Despite this co-occurrence, however, there are natural kinds that do not instantiate all the properties usually clustered, but just a subset of them, which is perfectly normal given the complexity of many natural phenomena and the "imperfection" of our scientific knowledge.

This theory can be useful once transferred into the debate about emergence. There is no reason to think that emergence can be captured by a fixed and definitive number of criteria once and for all because there is no reason to think that emergence is a univocal, unambiguous phenomenon that always occurs in the same way. The alleged presence of emergent phenomena in different domains of reality renders questionable the idea that in each of these different domains emergence exactly occurs in the same way. In fact, the plurality of definitions that can be found in different philosophical and scientific fields suggest exactly this: that emergence has different features according to the different contexts in which it takes place. This is the reason why I think that rather than being defined by an unchanging set of necessary and sufficient conditions, emergence can be correlated with an open cluster of properties including – at least – ontological and epistemological irreducibility, nonlinearity, fundamentality, qualitative novelty, and causal efficacy. This standpoint respects the pluralist instance so often highlighted by many scholars, and at the same time suggests what can be defined a "thick view" of emergence, opposed to the "sparse view" recognised by Mark Bedau and Paul Humphreys in the volume mentioned at the beginning of this introduction. While at the origin of the debate, emergence was correlated with facts that cannot be easily explained by science – the British Emergentists focused almost unanimously on the origin of life and mind – today emergence is recognised as a common trait of nature which does not represent an exceptional phenomenon, but rather the way in which matter is organised and become effective when structured.

[10] Boyd in Wilson 1999: 143.

CHAPTER I

Criteria for Emergence

1.1 *Taxonomies and meaningful criteria*

In the last few decades, the interest in the notion of emergence has grown steadily. As testified by several debates in philosophy and science,[1] emergence is alleged to be to be suitable to describe and better understand a conspicuous number of natural phenomena, such as, in physics,[2] spacetime,[3] quantum entanglement,[4] collective phenomena like phases of matter, superconductivity or ferromagnetism;[5] in chemistry,[6] covalent bonding,[7] and molecular macroscopic properties;[8] in complexity science,[9] stigmergy, flocking and similar coordinated behaviour of birds, fishes, and mammals, but also the development of urban centres,[10] and other social phenomena like the Web and urban centers.[11]

---

[1] See Gibb, Hendry, & Lancaster 2019.

[2] Laughlin and Pines (2000), Pines (2000) and Crowther (2016) talk about the emergence of complete new levels or domains of physics independent on lower levels, which they called, starting from Pines 2000, "quantum protectorates": "[…] quantum protectorates – stable states of matter whose generic low energy properties, insensitive to microscopics, are determined by a higher organizing principle and nothing else" (Pines, 2000: 341). In Pines opinion, for instance, the unpredictable phenomenal behaviour of superconductors falls in a quantum protectorate that renders in principle impossible its deducibility from the microscopic.

[3] Hu 2009; Butterfield 2011, Mattingly 2013, Bain, 2013, Crowther 2013 and 2016, Wüthrich, 2018.

[4] Humphreys 1997, Kronz and Tiehen 2002, Hüttemann 2005, Humphreys 2016, Silberstein 1999.

[5] Lebovitz 1999, Liu 1999, Zhang 2004, Laughlin 2005, Batterman 2001 and 2011, Pavarini, Koch, & Schollwöck 2013, Crowther 2016, Humphreys 2016.

[6] Hendry 2006, Manafu 2014, and for a list of irreducible chemical feature van Brakel, 2000.

[7] Humphreys 2016: 82 et seq.

[8] Luisi 2002, Scerri 2008, Humphreys 2016.

[9] Grassé 1959, Theraulaz & Bonabeau 1999, Cucker & Smale 2007.

[10] Stevenson 2002, Batty 2012, West 2017.

[11] Portugali 2011, Bretagnolle, Pumain, & Vacchiani-Marcuzz, 2009.

Despite being invoked in all these various scientific fields, however, no uncontroversial theoretical framework has been yet formulated, and there is no consensus in the scientific and philosophical world about the precise meanings, the methodological value, and the theoretical potentials of the idea of emergence.

In his book *Emergence. A Philosophical Account*, [12] Paul Humphreys examines the reasons of this complexity, which ascribes to two factors. The first is that "[…] the concept of emergence is not rooted within a single science",[13] therefore there is no discipline that could have the final say on it. The second reason is that "[…] we do not have a firm pre-theoretical grasp of emergence in the way we do with something like causation".[14] This second reason is important as it emphasises our inability to pre-theoretically agree on what is emergent and what it is not, and without uncontroversial cases of emergent phenomena it is hard to produce a corresponding uncontroversial, bottom-up definition of it.

Individual attitudes, as well as disciplinary origins, could not help, therefore, in providing a definitive and coherent description of emergent entities and behaviours; consequently, it may be better to consider the word "emergence" as an umbrella term encompassing different features and describing different phenomena, rather than a word indicating a precise natural process. As stated by Sophie Gibb, Robin Hendry and Tom Lancaster "[…] the philosophers and the scientists are using the same word to mean different things", [15] and this situation must be kept in mind when discussing about emergence.

Despite the clear pluralistic stance that seems reasonable to adopt,[16] however, it is still possible to distinguish certain criteria usually exploited to identify emergent phenomena, and this circumstance is manifest in three extensive studies on emergence provided in recent years by Paul Humphreys (2016), Carl Gillett (2016), and Jessica Wilson (2019). These three authors devoted special attention to the topic in metaphysics and philosophy of science, and wrote three ambitious monographs supposed to provide comprehensive accounts of emergence. The models they developed, as we will see, are different from each other, but at the right level of abstraction nonetheless three recurrent criteria for emergence can be recognised. The first is the *ontological irreducibility* of emergent phenomena:

[12] Humphreys 2016.
[13] *Ivi*: xvii
[14] Ibidem.
[15] Gibb, Hendry & Lancaster 2019: 2.
[16] See, again, Gibb, Hendry & Lancaster: "A less drastic response […] is pluralism: even if emergence is a somewhat amorphous notion, one might continue to use the term for importantly different kinds of phenomena while recognising that no single definition will do justice to all its uses" (3).

emergents cannot be decomposed in their smaller, more fundamental parts, without losing their essential nature in the process. The second is their *epistemological irreducibility*: emergents and their features cannot be exhaustively explained, predicted, or deduced starting from the knowledge of their base of emergence alone. The third criterion is *novelty*: emergents must either be fundamentally new, or exhibit new qualitative features, properties, and/or powers, or require new conceptual frameworks to be analysed.

Although characterized in different ways by the three authors mentioned above (and by the literature they examine), these three criteria are not only exploited by all of them, but also ubiquitous throughout the entire debate about emergence, where they are used, as we will see, for defining emergentists views in opposition to reductionism, physicalism, or epiphenomenalism.

Let's now start describing Humphreys', Gillett's and Wilson's account of emergence, and highlighting the kind of criteria they use, as well as their definition of them. Despite being omnipresent in the emergence debate, in fact, the meaning of the notions of "irreducibility" and "novelty" changes from an author to another, and this circumstance require certain theoretical clarifications necessary to fully understand what does it mean, for an emergent, to be irreducible or novel.

1.2 *Paul Humphreys*

Paul Humphreys' *Emergence. A Philosophical Account* shows that it is possible to distinguish different kinds of emergence according to the different ways in which what he dubs "Generative Atomism" (GA) fails. GA is described as a persistent view of nature for which entities are either fundamental "atoms" or non-atomic objects composed by the former in virtue of a "fixed set of rules that govern the construction process".[17] This atomism is called "generative" because non-atomic entities are *generated by* atomic ones and can also be decomposed again in the relevant fundamental entities. For these reasons, Humphreys states that GA guarantees the "in-principle predictability, the explainability and the lack of novel features of the whole system with respect to the properties of its parts".[18] Moreover, GA lead to "the reducibility of the whole system to its parts and hence to a lack of autonomy of the compound system".[19]

[17] Humphreys 2016: 12.
[18] *Ivi*: 13.
[19] Ibidem.

In particular, when GA's atoms are *physical* atoms, Humphreys talks about "GAP" (Generative Atomistic Physicalism), by which he means "an ontological position in which the fundamental entities are physical atoms […] and structured arrangement of these atoms determine the entire ontology".[20] In Humphreys view, this position is the basic theoretical frame of many physicalist theories such as David Lewis's Humean supervenience or David Armstrong's combinatorial ontology,[21] but it is quite reasonable to correlate GAP to microphysicalism as it is defined by Philip Pettit[22] or Andreas Hüttemann.[23]

In several cases, however, GA fails, and when this happens reductionism is ruled out and talk about emergence begins. In these cases, there are two ways in which it is possible to approach the question. First, by focusing on the type of *relationship* holding between the lower-level properties or entities that constitute the so-called "emergence base", and the higher-level "emergent" properties or entities arising from it. This relation can be ontological, pertaining the phenomena themselves, or inferential, or conceptual, pertaining our knowledge about them. Second, another approach to describe emergence emphasises the *temporal dimension* of emergent processes and distinguishes between synchronic and diachronic emergence. These outlined models of emergence are the following.

1.2.1 *Ontological emergence*

In the first case, emergent properties are genuine novel properties of the world. They are ontologically distinct from the properties from which they emerge and our knowledge about the system in which they appear is irrelevant to their existence. Ontological emergences seem to require clarifications concerning what kind of laws of nature regulate them and what novel features they introduce in reality.

The notion of *strong emergence*, which can be easily found in literature,[24] is a kind of ontological emergence implying the existence of novel, irreducible causal powers which allow for what is called "downward causation".

For instance, from O'Connor:

[20] *Ivi:* 17.
[21] *Ivi:* 18.
[22] Pettit 1993.
[23] Hüttemann 2004.
[24] See O'Connor 1994, Bedau 1997, Chalmers 2006 and Gillett 2016.

> Finally, there is the idea of "novel causal influence". This term is intended to capture a very strong sense in which an emergent's causal influence is irreducible to that of the micro-properties on which it supervenes: it bears its influence in a direct, "downward" fashion, in contrast to the operation of a simple structural macro-property, whose causal influence occurs via the activity of the micro-properties which constitute it.[25]

As we will see, Jessica Wilson agrees on this use of the term: "[…] strong emergentists maintain that some special science features are real, distinct, and distinctively efficacious as compared to their physically acceptable base features",[26] whereas Carl Gillett uses the expression "strong emergence" in a more restrictive way and does not identify strong emergence with a general ontological model, but with a particular case in which emergent properties are (i) physically realised and (ii) determinative.[27]

One of the central criteria for the attribution of the status of ontological emergence in Humphreys view is the presence of *novelty*, where novelty is an indicator of the failure of GAP. For Humphreys, an entity is novel "with respect to a domain D just in case it is not included in the closure of D under the closure criteria C that are appropriate for D".[28] In Humphreys view, therefore, novelty is always relative to an ontological domain; there is no *absolute* novelty, while there are different ways in which an entity can be novel. With respect to a particular ontological domain, for instance, an entity can be said novel if its relevant dependency relation – e.g. supervenience, or causation – with lower-level entities does not subsist. Another case in which novelty can appear is in the case of law closure: "[…] entities of type B are novel with respect to a domain D if and only if there is at least one law that applies to type B entities that does not apply to entities in D".[29] Type B entities, therefore, are not included in the ontological domain D under the relevant law closure criteria. An example of this kind of novelty comes from subatomic physics. Many elementary particles, such as muons or taus – which are types of not stable lepton – decay, or, as Humphreys prefers, "transform" in other types of more stable particles, such as neutrinos, quarks or antiquarks. With the appearing of these novel particles, new conservation laws appear too – or at least this is what the Standard Model implies. In

[25] O'Connor 1994: 106.
[26] Wilson forthcoming: 61.
[27] See § 1.3.3 below.
[28] Humphreys 2016: 29.
[29] *Ivi*: 32.

Humphreys' view, therefore, the emergent particles fail to be closed under the laws governing the particles from which they emerge, and this makes them novel because unable to satisfy GAP commitments.[30]

1.2.2 *Inferential emergence*

Inferential emergence is related to *unpredictability* and *underivability*. A clear way to frame this point is provided by Jaegwon Kim, who distinguishes between *inductive predictability* and *theoretical predictability,*[31] and states that even emergent phenomena can be inductively predicted, but they cannot be *deductively* predicted:

> What is being denied by emergentists is the theoretical predictability of [the emergent property] *E* on the basis of [the microstructural property] *M*: we may know all that can be known about *M* […] but this knowledge does not suffice to yield a prediction of *E*.[32]

Now, as a matter of fact, this idea was already present in John Stuart Mill and in British Emergentists. Writing about water in *A System of Logic*, for instance, Mill states that

> […] no experimentation on oxygen and hydrogen separately, no knowledge of their laws, could have enabled us deductively to infer that they would produce water. We require a specific experiment on the two combined.[33]

And the same can be found in George Henry Lewes' *Problems of Life and Mind*:

> Who, before experiment, could discern nitric acid in nitrogen and oxygen? Who could foresee that gold would be changed into a chloride if plunged into a mixture of two liquids (hydrochloric and nitric acid), in either of which separately it would remain unchanged?[34]

---

[30] *Ivi*: 66 et seq.
[31] Kim 1999: 8.
[32] Ibidem.
[33] Mill 1843: 255.
[34] Lewes 1977: 413.

In other words, we can *inductively* predict the appearance of an emergent phenomenon in a system thanks to the familiarity with its already known behaviour, but we cannot theoretically (*deductively*) predict it before its first appearance from the knowledge of its emergent base alone. In addition to this knowledge some experiments and an "inductive generalization"[35] will be required.

There are other variants of weak emergence, but it is rather uncontroversial that all of them entail a specific explanatory complexity: Gillett's account of weak emergence, for instance, is consistent with these ideas as well. In the case of Jessica Wilson, however, this correspondence between weak emergence and epistemological criteria does not hold, given that, as we will see, she draws the line between Strong and Weak schemas for emergence in metaphysical terms[36].

1.2.3 *Conceptual emergence*

This last model of emergence concerns properties that are defined as emergent because in order to explain them the elaboration of new conceptual frameworks is required. Every special science property is an example of – at least – conceptual emergence, for different domains need different theoretical frames to become tractable.

A famous and often quoted paper, about this idea, is Philip Anderson's *More is different*,[37] in which the author wrote: "Surely there are more levels of organization between human ethology and DNA than there are between DNA and quantum electrodynamics, and each level can require a whole new conceptual structure".[38] Anderson's point is pertinent. Over time, the number of the existing scientific disciplines, as well as that of the discovered and classified scientific entities and properties, has multiplied, rather than reduced, and this circumstance may indicate that the more we dig, the more – rather than the less – we need theories and models of reality.

[35] Humphreys 2016: 146.
[36] See § 1.4 below.
[37] Anderson 1972.
[38] *Ivi*: 396.

1.2.4 *Synchronic and diachronic emergence*

As far as temporal characterisations are concerned, Humphreys provides a distinction between *synchronic emergence,* in which higher-level and lower-level phenomena coexist at the same time, and *diachronic emergence*, in which the dynamics of the system is of primary importance as the new emergent properties diachronically appear from previous processes.

Given the particular attention payed to mental states as ideal cases of higher-level phenomena, philosophers and metaphysicians have generally focused on synchronic emergence, rather than diachronic one.[39] Synchronic emergence implies the coexistence of lower- and higher-level phenomena and poses some kind of synchronic dependence relations holding among them. In Humphreys' opinion, however, focusing on synchronic cases alone represents an obstacle in understanding the real potential and applicability of the ideas concerning emergence, because in many real systems, synchronic criteria alone cannot fully account for emergent phenomena. One of the examples used by Humphreys to demonstrate this point is quantum entangled states, where two or more original quantum systems interact and "transform" in a new, inseparable entangled compound whose state cannot be analysed in terms of its components alone (as well as the states of the different components, which becoming strongly correlated with one another, cannot be described independently). The model of ontological emergence Humphreys developed in his book is precisely meant to explain these kind of processes and it is, as a matter of fact, a diachronic model called "transformational emergence" which is a generalisation of the model he had described in his 1997 paper *How properties emerge*,[40] namely "fusion emergence".

Fusion emergence is diachronic because the involved emergent properties appear *after* certain dynamic processes or interactions of the parts of the system in which they appear, rather than in virtue of particular synchronic relations such as supervenience, realisation, mereological composition, and so on. A fusion process can be represented in the following way. $P_m^i\ (x_t)$ is the instantiation located in the space x at time t of the property Pm belonging to the domain i. Suppose to have another property instance Pn, belonging to the same domain i, and located in the same space x at time t: $P_n^i\ (x_t)$. After what Humphreys defines a "fusion interaction" indicated by the sign * (an intra-level operator), we will have a new fused property instance P_s belonging to the level j instantiated at time t' as follows:

[39] A significant exception to this trend is represented by philosophers of biology or philosophers of life sciences, who have been long focused on diachronic models of emergence.
[40] Humphreys 1997b.

$[\, P_m^i\ (x_t) * P_n^i\ (x_t)\,] \rightarrow P_s^j\ (x_{t'})$.

If the two levels i and j are different levels, the fused emergent property is a property of a new domain, and this is the most relevant fusion case, but fusion process do not necessarily imply a change of level.[41]

Fusion emergence has a significant advantage over synchronic accounts of emergence. The resulted property $P_s^j\ (x_{t'})$ cannot be analyzed in supervenient terms because it is a whole novel property and the original properties $P_m^i\ (x_t)$ and $P_n^i\ (x_t)$ producing it "no longer exist as separate entities"[42]; for this reason, this model of emergence avoids problems of causal overdetermination, because the original properties, being no more existent, cannot causally compete with the fused one. Humphreys' model of fusion emergence is therefore diachronic and, differently from synchronic models, does not imply an ontology constituted by levels.

These two features, diachronicity and what can be seen as a "flat" ontology, are the same Alexandre Guay and Olivier Sartenaer[43] highlighted in their account of "transformational emergence" (TE), which is explicitly inspired by Humphreys model.

TE can be described as follows:

> $E_{t'}^{l'}$ diachronically emerges on B_t^l (with $t' > t$ and $l' \geq l$) iff:
> (DEPd) B_t^l diachronically determines (*e.g.* causes) $E_{t'}^{l'}$; and yet
> (NOVd) $E_{t'}^{l'}$ is historically novel with regard to B_t^l.[44]

Here, we have an emergent entity E belonging to a level l' existing at time t' which emerges by transformation from an entity B belonging to the level l (which can be l' or a lower-level l'^{-1}) existing *before* the entity E, namely at time t (which is t'^{-1}). Guay and Sartenaer's model defines the dependency relation between the emergent entity E and its emerging base B as a diachronic one (DEPd), such as "spatiotemporally continuous [causal]

[41] This point is highlighted in Humphreys' 2016 book, because in the 1997 paper he assumed that fusion resulted in a change of domain, but this is a view he is not accepting anymore. Despite accepting a hierarchy of levels as a hypothesis, however, in 1997 Humphreys already considered this position "misleading and probably false" (Humphreys 1997: 5).
[42] *Ivi*: 5.
[43] Guay & Sartenaer 2016.
[44] *Ivi*: 298.

processes"[45] going from a previous state of the system S_1 to a following state S_2. As for novelty, the model assumes historical novelty (NOVd), rather than hierarchical, and correlates it to the appearance of new entities, properties or powers, governed by new laws.[46] Given this kind of novelty, if we associate the different states of the system S_1 and S_2 to the descriptive models M_1 and M_2 respectively, we can say that there is no way S_2 can be described by M_1. (DEPd) and (NOVd) imply, therefore, that "[…] prior to t_2, it is impossible in principle to predict or etiologically explain the nature and behavior of S_2 from complete knowledge of S_1".[47]

In both the cases examined above (and the same can be said for another recent account of diachronic emergence, namely Sartenaer's "flat emergence"[48]), it seems that diachronic accounts of emergence require the same two features highlighted in synchronic accounts: on the one hand, a partial dependence between the original and the emerging phenomena which render a smooth reduction (which in the diachronic case is a deduction rather than a reduction) of the latter to the former impossible, and, on the other hand, the appearance of some kind of novelty, which in this case is historical, which means, in C. Lloyd Morgan's words, that "there is more in the world to-day than there was in the primitive fire-mist".[49]

1.2.5 *Some remarks about this taxonomy*

I would like to highlight several points related to this taxonomy. First of all, differently from other taxonomies, which merely discern ontological from epistemological emergence, I think that the additional distinction of the second one between inferential and conceptual is an accurate choice, because inferential emergence just implies a negative "lack of derivability", while conceptual emergence focusses on the way in which we produce new, positive concepts to represent or describe emergents, such as when, for instance, we talk about the emergence of coordinated behaviour in birds and we have to introduce the notion of "flocking", or when we talk about the emergence of ordered states in phase transitions and we have to use the terms "liquidity" and "rigidity".

---

[45] *Ivi*: 303.
[46] See *ivi*: 304 "(NOVd) S2 exhibits new entities, properties or powers that do not exist in S1, and that are furthermore forbidden to exist in S1 according to the laws $\{L_1^i\}_{i=1}^{n}$ governing S1. Accordingly, different laws $\{L_2^i\}_{i=1}^{m}$ govern S2."
[47] *Ivi*: 307.
[48] Sartenaer 2017.
[49] Morgan 1913: 30.

Secondly, inferential and conceptual emergence are epistemologically characterised concepts, so individually taken they are metaphysically innocent, while, obviously, the first kind of emergence, the ontological one, it is not. The irreducibility and unpredictability of ontological emergent phenomena are, therefore, *in principle* limitations, while they are at least *in practice* limitations as far as epistemological emergence is concerned. If we suppose that inferential emergence is a purely in practice limitation on deducibility, a hypothetical "laplacian calculator"[50] would predict perfectly well every possible and actual future behaviour of the system, as the actual difficulty in prediction would merely be due to the backwardness and insufficiency of our present calculation tools. However, ontological and epistemological irreducibility should not be considered as distinct features. In the original formulation of the concept of emergence, in fact, epistemological irreducibility was just the consequence of ontological irreducibility, and this is the reason why some British Emergentists like Broad[51] and Alexander[52] would be more the ready to abandon the notion of emergence if an exhaustive scientific explanation of the phenomena at issue would be available.

The category of emergent phenomenon, in other words, can be said to be really interesting when the emergence in question is metaphysical and the epistemological emergence is a consequence of it. "Interesting", however, does not mean "real" or "existing": it is not my intention to downgrade epistemologically emergent phenomena to phenomena that are only *apparently* emergent. Precisely because of the pluralist position that I embraced from the very first pages of the introduction, I think it is more than reasonable to consider these phenomena as emergent too, but the difference between them and ontologically emergent phenomena must be duly taken into account and will be illustrated in detail in the

[50] In his *Seeking Ultimates. An Intuitive Guide to Physics*, the theoretical physicist Peter T. Landsberg describes the laplacian calculator as "[…] a dematerialized intelligence, a kind of God who knows of all collisions, can distinguish all microstates in a fine-grained phase space, and all his calculations of future and past states (in so far as allowed by science) are performed instantaneously" (Landsberg 1999: 86).

[51] See Broad 1923: 55 "It was held that the characteristic differences between the behavior of Oxygen and Hydrogen are due in no way to differences of structure or components, but must simply be accepted as ultimate facts. This first alternative can hardly be counted as one way of explaining differences of behavior, since it consists in holding that there are certain differences which cannot be explained, even in part, but must simply be swallowed whole with that philosophic jam which Professor Alexander calls "natural piety". It is worthwhile to remark that *we could never be logically compelled to hold this view* [emphasis mine]".

[52] The famous passage from Alexander in which he states that emergents should be accepted "under the compulsion of brute empirical facts" (1920: 47) means that we have to accept emergence given the empirical data we have. But Alexander also confess his "feeling, as a metaphysician, a horror of notions which the mind takes for ultimate and undefinable". Emergent phenomena have to be accepted, therefore, given the actual available knowledge, but nothing prevents a revision of the explanation of these entities in the future. As pointed out by Symons (in Vintiadis & Mekios 2018: 181) "Alexander and Broad acknowledge the possibility that apparent differences are illusory and can be explained away with the progress of inquiry".

Conclusions, when I will use the image of the *Target* to define how to conceptualize these different types of emergences.

Third, despite a certain diffused parsimonious attitude which Bedau and Humphreys had defined "the sparse view",[53] and Humphreys renames the "rarity heuristic",[54] it is worth noting that the idea that authentic emergent phenomena are rare and uncommon in nature – or even occurrent in the mental realm alone[55] – is no longer dominant in the philosophical and scientific debate.[56] This idea was ancient, given that even John Stuart Mill, talking about the way in which causes and effect can be jointly composed, considered heteropathic effects – those which will be considered emergent – rare: "The former case, that of the Composition of Causes, is the general one; the other [the heterogeneous composition] is always *special* and *exceptional* [emphasis mine]".[57] In spite of this general, parsimonous idea, it seems that acceptable cases of emergence are more common than expected, even in physics.[58] Such a pervasiveness of emergent phenomena in the sciences provides a further suggestion. Emergence processes are widespread and they do not concern merely cryptic phenomena like life or consciousness. Therefore, rather than imposing a unique definition of emergence in a metaphysically prescriptive way, it could be more reasonable taking into account all these variants of emergence and adopting a more pluralistic position, aware that different contexts require different conceptions.

The pluralistic view was first highlighted by Mark Bedau, who writes that

> […] each conception of emergence must be evaluated independently and on its own merits. It is not competing against other conceptions for the top spot on the podium. Every viable conception would share space on the podium with the other leading candidates for actual and important kinds of emergence.[59]

[53] Bedau & Humphreys 2008: 12.
[54] See Humphreys 2016a and 2016b: "[…] the rarity heuristic: any account of emergence that makes emergence a common phenomenon has failed to capture what is central to emergence. Those early twentieth-century writers who restricted emergence to phenomena that at the time seemed mysterious and little understood, such as life and consciousness, seemed to have been sympathetic to the rarity heuristic, although earlier philosophers, such as Mill, who considered chemical properties to be emergent, would have rejected it" (Humphreys, 2016a: 760).
[55] See Newman 1996, McLaughlin 1997, McIntyre 1998, Kim 1999, Chalmers 2006.
[56] See Bedau & Humphreys 2008: 12 et seq. Humphreys 1997: 54 et seq.
[57] Mill 1843: 373.
[58] About this, see the examples I made in § 1.1.
[59] Bedau 2013: 92.

As we mentioned at the beginning of the chapter, however, this awareness is now more common, as recently testified by the authors of the *Routledge Handbook for Emergence* published in 2019, who suggest that "even if emergence is a somewhat amorphous notion, one might continue to use the term for importantly different kinds of phenomena while recognising that no single definition will do justice to all its uses".[60]

Another central point to emphasise, the fourth, is that the ubiquitous use of the term in science seems to imply that there are at least some notions of emergence that are not at all unscientific; more unscientific seems, rather, the desire to determine its definition once and for all, out of an appropriate context. Related to this, it is worth mentioning that some literature reports two approaches to questions such as emergence. On the one hand, there is the so-called "science-first" approach in which metaphysics and its problems are driven by scientific knowledge: "This 'science-first approach' […] to emergence and reduction holds that metaphysics should be subsumed under first-order philosophy of science as much as possible, for example, philosophy of physics, biology, cognitive science, neuroscience, etc".[61] About this, Humphreys himself highlights the importance of "[…] a metaphysics of emergence informed by science, rather than by pure apriori analysis".[62] On the other hand, this latter "apriori analysis" can be called a "philosophy-first" approach, and focuses on developing abstract and general models supposed to clarify the scientific debate. Carl Gillett, for instance, talking about emergence but also about reduction, claims that

> […] philosophers have erected their own proprietary views of the nature of scientific composition, reduction, and emergence *diverging from the accounts apparently used in the sciences* [emphasis mine]. Applying their different theoretical frameworks, the reigning view in mainstream philosophy is consequently that reductionism is basically a *dead*, and perhaps even somewhat *distasteful*, position. And many (most?) philosophers dismiss discussions of emergence as, at best, *kooky* and, at worst, *incoherent* (to use far more polite terms than are usual in such dismissals). [63]

[60] Gibb, Hendry & Lancaster 2019: 3.
[61] Silberstein 2012: 640. See also Silberstein 2011 and 2012, and Crowther 2016.
[62] Humphreys 2016: 15.
[63] Gillett 2016: 4-5.

As suggested by Gillett, the second approach, which he defines as "metaphysics *for* science",[64] rather than "metaphysics *of* science", gives rise to several complications, and, frequently, provides inappropriate theoretical frameworks for scientific concepts which are used in philosophical debates specifically for their scientific value – even if this value, during their "migration" from science to philosophy, is often lost.

Connected to these questions, there is an additional fifth point that I mention here and I will develop in next paragraphs and chapters. The widespread use of the concept of emergence in science suggests not only that emergence can be a scientifically useful and legitimate concept, but also that it should not be stubbornly opposed to all forms of reduction, which is an essential scientific theoretical tool. As a matter of fact, as Wimsatt and Sarkar[65] rightfully noticed, "Within practical scientific contexts, emergence and reduction are not usually regarded as opposites".

1.3 *Carl Gillett*

While Paul Humphreys correlated emergence to the failure of GA, and consequently considered irreducibility a necessary condition for ontological emergence, in Carl Gillett's view, once clarified the meaning emergence and reduction, there is no conflict between these two notions. We said that Gillett's taxonomy is built on a more scientific background which counterbalances his manifest suspiciousness towards too speculative philosophical accounts of emergence and reduction. On the heels of William Wimsatt[66] and John Bickle,[67] Gillett strongly expresses his lack of confidence in speculative philosophy, and highlights the necessity to focus on science first. It is worth noticing, moreover, that Gillett's account of emergence is the reformulation and development of Samuel Alexander's theories, a scholar who studied and taught scientific psychology and worked for years in the physiology laboratories in Oxford. Therefore, as pointed out by Gillett, "this scientific activity partly explains Alexander's commitment to a metaphysics heavily informed by the findings of the sciences".[68]

---

[64] See Gillett 2016: "[…] my overview of various philosophical debates, whether about reduction or scientific composition, suggests philosophers have to different degrees fallen into what I term "metaphysics *for* science" in the practice of unwittingly shoehorning scientific concepts and positions into unsuitable theoretical machinery developed for other purposes" p. 11.
[65] Sarker 2006: 702
[66] Wimsatt 1976.
[67] Bickle 2008.
[68] Gillett 2006: 265.

Now, in Gillett's opinion, the root of the philosophical misunderstanding of the notion of emergence lies in an incorrect consideration of the elementary idea of "qualitative emergence". For Gillett, every physical aggregative compound manifests, in some sense, novel features and behaviours, but this circumstance does not really entail the failure of reductionism, as well as the authentic emergence of novel phenomena. For Gillett, qualitative emergence is merely the fact that in case of aggregates, wholes have properties not had by parts, but this condition, he states, is not inconsistent with reductionism.

Let me briefly illustrate, now, the three different forms of emergence described by Gillett. He uses two crucial notions, those of realisation and determination/production, which I will define in the next paragraphs. For now, we can just consider that realisation means some sort of synchronic property composition, and determination/production a strict notion of causal efficacy. A detailed description will follow.

1.3.1 *Weak Emergence (or W-Emergence)*

In W-Emergence, a higher-level property *X* is W-emergent if (i) it is *realised* by the property of the lower-level individual from which it emerges, and (ii) its relevant statements, laws, explanations, and theories cannot be *deduced*, *predicted* or *computed* by the statements, laws, explanations, and theories defining the lower-level individual.

Developed in the framework of complexity sciences, this model of emergence, as well as the weak emergence described by Humphreys, is characterised by the notions of underivability and unpredictability, and it is, in fact, metaphysically innocent. It admits the ontological realisation/composition of higher-level properties, and leaves room for a semantic and epistemic kind of autonomy for higher-level sciences and theories.

1.3.2 *Ontological Emergence (or O-Emergence)*

In O-Emergence, a higher-level property *X* is O-emergent if (i) it is an *unrealised* property instance and (ii) it is "productive and hence determinative".[69]

Ascribed by Gillett to several theories of mind (in particular theories of phenomenal consciousness), this model of emergence is robustly ontological, but seems scientifically controversial, as it adopts an unrealised view of O-emergent phenomena.

[69] Gillett 2016: 183.

As we will see, in Gillett's view unrealisability is a deeply controversial feature, whereas, by contrast, realisation is hardly dispensable in the scientific framework because it is grounded on compositional explanation, and entails, as a consequence, a compositional view of reality.

1.3.3 *Strong Emergence (or S-Emergence)*

In S-Emergence, a higher-level property *X* is S-emergent if it is (i) *realised* by the properties of the lower-level individual from which it emerges and (ii) it is *determinative*.

This further model of reduction is attributed by Gillett to scientists such as Philip Warren Anderson and Robert Laughlin, and represents, in his opinion, a promising *via media* which Samuel Alexander already formulated in his *Space, Time and Deity.*[70]

S-Emergence is a metaphysical account that, on the one hand, can be compatible with certain reductionist suggestions, while, on the other hand, can empirically question a global, radical view of reduction. The problem with this model of emergence, however, is that realisability and determinativity are considered incompatible by Gillett.

1.3.4 *Composition, realisation and determination*

While Humphreys' twofold characterisation of epistemic emergence (inferential *and* conceptual) collapses in Gillet's single category of W-emergence, his single ontological account of emergence doubles in Gillett's O-emergence and S-emergence.

[70] In his paper dedicated to Alexander, Gillett describes his theory of emergence in these terms: "This is the heart of Alexander's concept of an 'emergent' property, as I shall further illuminate below: A realized property instance that partially, non-causally determines some of the contributions of causal powers of its realizers. Although Alexander never explicitly defines his notion of emergence we can thus frame his notion more precisely as follows: A property instance X is emergent, in an individual s, if and only if (i) X is realized by microphysical properties/relations, and (ii) X partially non-causally deter mines some of the causal powers contributed by at least one of the microphysical properties/relations realizing X" (272). In the next paragraphs, we will see how Gillett's account of strong emergence reflect Alexander's account.

One could say that while Humphrey develops a general ontological account, and two more precise epistemological accounts, Gillett does the opposite, providing a general epistemological account and two ontological versions presenting two different ways in which emergent entities could be ontologically determinative.

At this point, it is essential to define what do realisation, productivity, and determination mean, in order to understand the real connotation of Gillett's definitions. We will start from realisation, but before that, it is worth noting that, in Gillett, realisation is strictly connected with scientific composition, which in my understanding is one of the keystone of the reduction/emergence debate and deserves a primary analysis.

1.3.4.1 *Scientific composition*

Scientific explanations are compositional*,* which means that science explains the nature of higher-level phenomena through the nature of the qualitatively different lower-level phenomena composing them. This relation of composition is characterised by Gillett as follows. Composition is a vertical, metaphysical relation. It is, following his example, the relation holding between a diamond and its composing carbon atoms. This example shed light on the fact that it is a many-one constitutive relation (many carbon atoms constitute a diamond), it is asymmetric (the carbon atoms compose the diamond and not *vice versa*) and irreflexive (the carbon atoms do not compose themselves).

	Paul Humphreys	**Carl Gillett**
Ontological models	Ontological emergence	Ontological emergence
		Strong emergence
Epistemological models	Inferential emergence	Weak emergence
	Conceptual emergence	

Figure 1.1. Comparison between Paul Humphreys' and Carl Gillett's taxonomies of Ontological and epistemological models of emergence.

Composition is a synchronic relation holding between distinct entities, and it is therefore different from causation, which seems to be, moreover, diachronic, and from identity. Its holding, furthermore, is due to nomological necessity and depends upon particular background conditions. Eventually, compositional relationships concern different kinds of entities, for which, following Gillett, it is useful to adopt different terms. We will have, therefore, *part-whole relations* between individuals, *realisations* between properties, *implementations* between processes, and *comprisings* between powers. All these relations can be subsumed under the category of "compositional relation".

1.3.4.2 *Realisation*

Realisation, in Gillett's framework, is therefore the compositional relation holding between different properties, instantiated in different individuals which bear a particular spatiotemporal and functional organisation. This is the case, indeed, of W- and S-emergence, where some lower-level properties *F1…Fn* instantiated in lower-level individuals *s1…sm*, jointly realise the higher-level property *G*, instantiated in the higher-level individual *s**. This process, together with all the relevant compositional relations between powers and processes which accompany property realisation is described as follows:

> (Realization – JRF[71]) Property instances F1–Fn, in individuals *s*1–*s*m realize a property instance G, in individual *s** under background conditions $, *if and only if*, under $, (a) *s*1–*s*m are members of, or are identical to, a group of individuals *s*1–*s*n spatially contained within s*, (b) *s*1–*s*m bear spatiotemporal, productive, and/or powerful relation to one another, (c) *s*1–*s*n through their joint productive role-filling together non-productively result in *s** under $, but not vice versa, (d) the powers contributed by F1–Fn to *s*1–*s*m together through their joint productive role-filling non-productively result in the powers individuative of G, in s* under $, but not vice versa, and (e) the processes based by F1–Fn under $ are or would jointly non-productively result in all the processes that are or would be based by G under $ but not vice versa.[72]

[71] JRF means "joint role-filling". See footnote 73.
[72] Gillett 2016: 89.

To untangle these technicalities, imagine again a diamond. Carbon atoms are the lower-level individuals *s*1...*s*m which compose the higher-level individual *s** (the diamond). The atoms *s*1...*s*m have the properties F1…Fn (in particular: their chemical bonding and alignment), which together realise the higher-level property G (diamond's hardness), through a kind of many-one relation Gillett defines "joint role-filling".[73] These components *s*1...*s*m bear furthermore a particular spatial and functional organisation, where their lower-level powers and processes compose the correspondent diamond-level powers and processes.

Now, if W- and S-emergence concern realised properties, following the just mentioned schema, O-emergent properties are supposed to be *unrealised*, where unrealisability means having no components, and having no components, in a traditional sense, is equivalent to being fundamental. Physics, nevertheless, states that there are only four existing fundamental forces, i.e. the gravitational, the electromagnetic, and the weak and strong nuclear forces, therefore it seems that admitting new fundamental properties or powers can be scientifically problematic. This is the reason why Gillett liquidates this form of emergence: because it is an account that cannot be consistent with everyday scientific practice, or, as he calls it, with "everyday reductionism".

However, the other two forms of emergence described by Gillett present problems as well. As for the first, W-emergence, it is an epistemic account which does not provide any information about the ontological nature of the world. As for the second, S-emergence, it seems that realisation and determination, as we said, are incompatible. At this stage, the point is what does determinative mean and why being "determinative" seems excluded by being realised.

1.3.4.3 *Determination/production*

Gillett uses the word "determinative" meaning the feature of several properties which contribute ("make a difference"[74]) to the powers of the individuals in which are instantiated. The word "production" means the same in the case of processes. In sum, a property is determinative if it makes a difference to the individual, while a process is productive if it

[73] Gillett 2016: 359 "Joint role-filling – A many–one relation between working entities where none of the relata plays the role individuative of the other relata, but where these relata have roles that together fill the role of the other entity, thus entailing that the relata of joint role-filling are qualitatively distinct".
[74] See Gillett 2016: 64.

causes some effects that make a difference to the powers of the individual. The reason why determination could not be compatible with realisation descend from a strict notion of aggregation usually connected with composition and realisation. Gillett defines this view the "Simple View of Aggregation",[75] and states that the real theoretical disagreement between reductionists and emergentists lies on the very nature of aggregation. The Simple View, which is, for Gillett, "[…] an ontological cornerstone of the dominant strain of scientific reductionism",[76] consists in two theses. Firstly, it states that higher-level entities are *composed by* parts that are determinative, but the collective whole composed by them should not be attributed of an own determinative power for reasons of parsimony.[77] The only determinative entities are, therefore, the composing ones, and the only determinative relations are those holding between the lower-level composers. Secondly, the Simple View of Aggregation entails that in nature there is no discontinuity at different levels of organisation, because if the only determinative entities are the lower-level ones, these powers are inter-level and rule across all scales of collectives. However, this view, which Gillett defines "Simple Fundamentalism", is debateable, because as science clearly demonstrates, individuals behave in different ways depending on whether they are isolated, in simple systems, or in complex ones: "Parts behave differently in wholes".[78] For this reason, Gillett describes an alternative view of aggregation, the "Conditioned" view, which ascribes to Robert Laughlin. Gillett states:

> […] Laughlin is suggesting that our empirical findings show that certain components sometimes contribute different powers, and hence behave differently, under the condition of composing a certain higher-level entity, but where the component would not contribute these powers if the laws applying in simpler collectives exhausted the laws applying in the complex collective.[79]

The continuity of reality is, therefore, the weakest assumption of the Simple view of aggregation, because nature shows several cases of what Laughlin defines "insensitivity to microscopic" [80] or "walls of scales". [81] Several higher-level phenomena such as

[75] *Ivi*: 112.
[76] *Ivi*: 194.
[77] Ibidem.
[78] *Ivi*: 195.
[79] *Ivi*: 194.
[80] See Laughlin in Bedau & Humphreys 2008: 261.
[81] Ibidem.

superconductivity, superfluidity, or ferromagnetism, for instance, exhibit independence on microscopic details, creating what we can call "determinative (and consequently explanatory) gaps". These issues are connected to the idea of quantum protectorate,[82] namely "a stable state of matter, whose generic low energy properties are determined by a higher-organizing principle and nothing else".[83]

In sum, the possibility of a realised but still determinative composed entity, which is what Gillett's favourite account of emergence (S-Emergence) suggests, seems to be ruled out by a notion of aggregation that is no longer appropriate. Therefore, Gillett urges to reconsider the concept of aggregativity abandoning the Simple account and adopting an alternative view, the Conditioned one, which allows for a novel form of realisation allowing, in its own turn, for a novel form of determination Gillett calls "machresis" or "foundational determinative relation" (FDR).[84]

Machretic determination, or FDR, is a downward determination relationship holding between the composed emergent entity and its components, and it finds room in the worldview Gillett calls Mutualism and opposes to Fundamentalism. The latter position claims that all existing entities are either fundamental microphysical entities, or macroscopic entities composed by them, and to this claim, it adds (for reasons of parsimony, as we saw) that the only determinative entities are the fundamental ones. By contrast, the Mutualist view, which is the one supported by scientific emergentists such as Laughling,[85] implies that higher-level macroscopic entities, despite upwardly composed by the lower-level ones, are also downwardly machretically determinative upon them. A way to express this idea is by using the words of the philosopher of biology Charles Dyke, who talks about "structured structuring structures".[86] Strongly emergent entities are "structured" because they are composed, but they are also "structuring" in machretically constraining or shaping the roles of their composers. In this Mutualist view, entities are subject to a double vector of determination and are mutually interdependent upon one another. The two determination relationships, however, are of different kinds, for in the case of upward composition we have a compositional *role-filling* relation, while in the case of downward machresis, we

[82] See Laughlin and Pines 2000, Anderson 2000, Laughlin 2008.
[83] Laughlin and Pines 2000: 29.
[84] See Gillett in Gibb, Hendry & Landcaster 2019.
[85] Other authors Gillett defines as scientific emergentists are Philip Anderson, Iain D. Couzin, Jens Krause, Chris Langton, and Walter Freeman.
[86] Dyke 1988: 24.

have a non-compositional and "non-causal"[87] *role-shaping*. Machresis, therefore, is a *sui generis* determinative relation which constrains the lower-level parts of a whole to have some differential powers, but is compatible with composition. This compatibility, eventually, is what renders S-emergence scientifically acceptable, for S-emergent phenomena are both compositionally constituted and machretically determinative: they are aggregates composed by lower-level parts analysable through compositional explanations, but they are also wholes able to "make some difference" to their parts and to the other individuals at their own level.

In this framework, emergent phenomena are irreducible to their parts because the individual properties of the parts are not the only properties that the emergent entities exhibit. The emergents, on the one hand, exhibit constraints on lower-level parts in the form of machretic determination, and, on the other hand, being mutually interdependent with their parts, they jointly contribute to causing effects both horizontally at their own level and "diagonally" in a diachronic intra-level way.

In conclusion, as well as Humphreys, who describes a Generative Atomistic Physicalism and correlates emergence with its failure, Gillett portrays a Fundamentalist physicalism characterized by a particular, Simple view of aggregation – which is associated with reductionism – and notices that it is appropriate to admit S-emergent entities in the several cases in which this simple view cannot explain the phenomena at issue.

1.4 *Jessica Wilson*

Jessica Wilson's *Metaphysical Emergence* is devoted, as the title suggests, to an analysis of a particular kind of emergence, namely the ontological or – as she defines it – *metaphysical* one. Wilson takes account of macroscopic special science entities and attributes to them two characteristics. Special science entities *depend upon* certain complex configurations of smaller, more fundamental entities, being *synchronically materially composed* by them; in a similar way, special science features are "at least partly"[88] determined by the features of these "micro-configurations", as she defines them. However,

[87] In his book, Gillett defines machresis as "non-causal", given that machresis does not involve any *activity*. He later reconsidered this definition, however, which he judges as not particularly fortunate (personal communication). We agree on that, and suggest that rather than being non-causal, machresis, as well as other forms of determination, are causal relationships able to produce changes in reality even if they do not involve physical processes. To admit these forms of causation, however, the notion of causal efficacy should be widened, and this is the suggestion we formulate in the last chapter of this work.
[88] Wilson forthcoming: 1.

	Paul Humphreys	**Carl Gillett**	**Jessica Wilson**
Ontological models	Ontological emergence	Ontological emergence	Strong emergence
		Strong emergence	Weak emergence
Epistemological models	Inferential emergence Conceptual emergence	Weak emergence	

Figure 1.2. Comparison between Paul Humphreys' and Carl Gillett's and Jessica Wilson's taxonomies of emergence.

in spite of the above, special science entities also exhibit a degree of ontological and causal autonomy, being them "[…] distinct from and distinctively efficacious as compared to the micro-configurations upon which they depend".[89]

Special science entities, in short, present (i) a *synchronic material dependence* on micro-configurations, and (ii) an *ontological and causal autonomy* from them. For Wilson, the coupling of (i) and (ii) defines emergence, and emergence is metaphysical because (i) and (ii) represent real features of entities themselves, and not some difficulties in our understanding, measuring, or representing them.

Before sketching Wilson's schemas for emergence, however, it should be recalled that both Humphreys and Gillett draw a distinction between ontological and epistemological accounts and this distinction is essentially shared by the vast majority of philosophers,[90] who talk about Weak and Strong emergence as well, defining the first as an ontological model, and the second as an epistemological one. Although Wilson too uses this very same terminology, it is worth noticing that she draws the line between weak and strong emergence in metaphysical terms alone, so her taxonomy presents two models of emergence that are both metaphysical.

1.4.1 *Two schemas for emergence*

Wilson poses two key questions in her book: what is emergence, and whether there are real cases of emergence in nature. To answer to these questions, she describes two schemas

[89] *Ivi*: 2.
[90] See, for instance, Chalmers 2006, Bedau 1997, O'Conner 1994.

for metaphysical emergence and presents them as the only viable models to which, significantly, almost every other accounts found in literature can be reduced. Wilson's two models of emergence depend upon the satisfaction of two conditions. The first is called the *New Power Condition*, while the second is the *Proper Subset of Powers Condition*. The fulfilment of the first one allows for Strong emergence, while the fulfilment of the second one allows for Weak emergence. Before considering what the conditions declare, however, some remarks are required.

First, Wilson talks about higher-level and lower-level features meaning by the latter the features of special sciences entities which are less fundamental than the features of physics. When *new* powers are mentioned, moreover, the novelty at issue is fundamental novelty[91].

Second, Wilson's conditions and schemas focus on emergent features and powers rather than on emergent entities because of the following idea. In agreement with Mark Bedau[92], Wilson states that an entity is emergent if it instantiates emergent features, and these features are emergent in their own turn because of the emergent character of their powers. Therefore, powers are central in Wilson's discourse for every emergent phenomenon can be traced back to emergent features and then, ultimately, to emergent powers.

Third, Wilson mainly talks about *token* features and powers, rather than *type* features and powers. The reason for this is that she adopts a "metaphysically neutral account" of powers, merely meaning, by them, "[…] what causal contributions possession of a given feature makes (or can make, relative to the same laws of nature) to an entity's bringing about an effect, when in certain circumstances".[93] Given that causation is a relation holding between spatiotemporally located phenomena, it follows that talk of powers implies talk of spatiotemporally located token of certain type powers, rather than talk of type powers themselves.

Now, let's return to the aforementioned schemas for emergence.

1.4.1.1 *Strong emergence*

The New Power Condition states the following.

[91] See Wilson 2016: 61.
[92] See Bedau 2002.
[93] Wilson 2019: 46.

> *New Power Condition*: Token higher-level feature *S* has, on a given occasion, at least one token power not identical with any token power of the token lower-level feature *P* on which *S*, on that occasion, synchronically materially depends.[94]

In this case, to fulfil the condition, it is necessary that the higher-level entity having a higher-level feature *S* has at least one power not had by the feature *P* of the lower-level entity on which the higher-level one materially depends. If this feature *S* has this *new* power, then that feature can be considered Strongly metaphysically emergent, rendering Strongly metaphysically emergent the entity itself.

The point to clarify, here, is how the fulfilment of the New Power Condition leads to Strong emergence. The answer is that an entity instantiating a feature having a new fundamental power cannot (by Leibniz's law) be identical to an entity which does not instantiate that feature and which, as a consequence, does not exert that very same power. This argument leads, therefore, to the ontological autonomy of the entity at issue. As for the causal autonomy, the argument is quite the same. The higher-level entity cannot have causal powers identical to those of the lower-level entity because the former has different features which have, in turn, different powers. Being therefore both ontologically and causally distinct because of the presence of a new power, the entity fulfilling the New Power condition results Strongly metaphysically emergent.

In Wilson's words:

> *Strong emergence*: Token apparently higher-level feature *S* is Strongly metaphysically emergent from token lower-level feature *P* on a given occasion just in case, on that occasion, (i) *S* synchronically materially depends on *P*, and (ii) *S* has at least one token power not identical with any token power of *P*.[95]

1.4.1.2 *Weak emergence*

Let's turn to the second case. The Proper Subset of Powers Condition states the following.

[94] *Ivi*: 64.
[95] *Ivi*: 67.

> *Proper Subset of Powers Condition*: Token higher-level feature *S* has, on a given occasion, a non-empty proper subset of the token powers of the token lower-level feature *P* on which *S* synchronically materially depends, on that occasion.[96]

To fulfil the condition, it is necessary that the higher-level entity having a higher-level feature *S* has a proper subset of the powers had by the feature of the lower-level entity on which the higher-level one materially depends. If the feature at issue has this proper subset of powers, then the feature can be considered Weakly metaphysically emergent, rendering Weakly metaphysically emergent the entity instantiating it.

> *Weak emergence*: Token apparently higher-level feature *S* is Weakly metaphysically emergent from token lower-level feature *P* on a given occasion just in case, on that occasion, (i) *S* synchronically materially depends on *P*; and (ii) S has a non-empty proper subset of the token powers had by *P*.[97]

Similarly to the case of the New Power Condition, the fulfilment of the Proper Subset Condition entails both ontological and causal distinctness of the higher-level entity. Having different sets of powers, the higher-level and the lower-level entities will be different by Leibniz's law (ontological distinctness) and will produce different effects (causal distinctness due to different causal profiles[98]).

For Wilson, therefore, it is possible to save the distinctness and causal efficacy of special science entities having both novel causal powers – as in the case of the fulfilment of the New Power Condition – or "a distinctive set (collection, plurality) of powers"[99] – as in this second case.

1.4.2 *The problem of the higher-level causation*

In defining emergence, Wilson gave a particular emphasis to causal powers. This circumstance has a precise theoretical motivation, which is the urge to overcome what she defines as the problem of higher-level causation, which, since her first formulations,

[96] *Ivi*: 71.
[97] *Ivi*: 85.
[98] *Ivi*: 79.
[99] Ibidem.

describes as “the primary challenge to the claim that higher-level entities and features may be metaphysically emergent”.[100]

The problem of the higher-level causation, also known as the *overdetermination* or the *exclusion* problem,[101] lies in the apparent impossibility, for a higher-level entity, to be distinctively efficacious in a world in which every (physical) effect is supposed to be produced by an equally physical (but lower-level) cause. If in this framework another cause is admitted, it follows that the same effect has two sufficient causes, leading to a case of causal *overdetermination*.

For Wilson, this problem can be exhaustively described listing six premises: on the one side, there are four features of higher-level entities – dependence, reality, efficacy and distinction (1-4) – and, on the other side, two principles concerning the supposed nature of causation – the Causal Closure of the Physical World and the Non-overdetermination requirement (5-6). As we will see, the acceptance of the four features renders impossible the contemporary commitment to both of the two principles. The six premises leading to the problem are following:

> (1) *Dependence*. Special science features synchronically materially depend on lower-level physically acceptable features […].
>
> (2) *Reality*. Both special science features and their physically acceptable base features are real.
>
> (3) *Efficacy*. Special science features are causally efficacious.
>
> (4) *Distinctness*. Special science features are distinct from their base features. […]
>
> (5) *Physical Causal Closure*. Every lower-level physically acceptable effect has a purely lower-level physically acceptable cause. […]
>
> (6) *Non-overdetermination*. With the exception of double-rock-throw cases, effects are not causally overdetermined by distinct individually sufficient synchronic causes.[102]

As mentioned, accepting the dependence, reality, efficacy, and distinctness of special science entities implies the failure of one of the other two premises, and the same can be said about the commitment to the last two premises. If both the Physical Causal Closure and the Non-overdetermination premises are accepted, at least one of the features of special

[100] Wilson 2014: 351.
[101] Wilson relates to Kim’s and Merricks’s formulations of the argument. See Kim 1993 and Merricks 2003.
[102] Wilson forthcoming: 55-56.

science entities listed above is to be rejected. To clarify this point, Wilson uses two examples.

First, suppose to have a case of *intra-level* causation between special science features having the four characters listed above (1-4). Special science feature *S* causes special science feature *S**, which is materially dependent on a lower-level feature *P** that, at least nomologically, necessitates *S**. For the Physical Causal Closure (5), however, *P**, to be produced, must have a purely physical cause, i.e. *P*, which is sufficient for the instantiation of both *P** and, by the instantiation of *P**, *S**.

*S**, therefore, seems causally overdetermined (failure of 6) by both S and P.

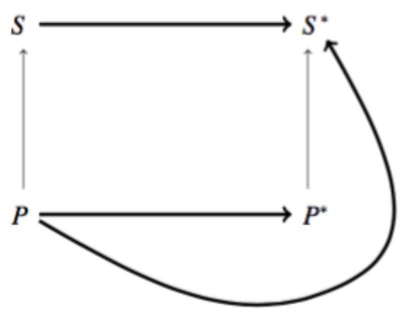

Figure 1.3. S causes S*, which seems caused by P as well

The second example is a case of *inter-level* causation between a special science feature having, again, the four characters listed above (1-4) and a lower-level, more fundamental feature. Special science entity S causes a lower-level basic feature *P**. For the Physical Causal Closure (5), although, *P** must have, in order to be produced, a purely physical cause, i.e. *P*, which is sufficient for the instantiation of *P**.

*P**, therefore, seems causally overdetermined (failure of 6) by both *S* and *P*.

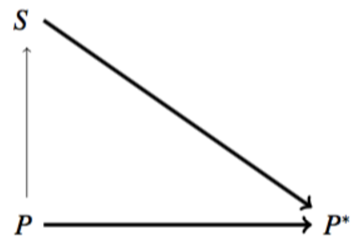

Figure 1.4. S causes P*, which seems caused by P as well.

To overcome the problem of the higher-level causation there are different strategies, each of them coinciding with the rejection of one or more premises of the list: in Wilson's opinion, dualism rejects dependence, eliminativism reality, epiphenomenalism efficacy, and reductive physicalism distinctness. All these strategies succeed in preserving the Physical Causal Closure and the Non-overdetermination principles, but at the same time they weaken special science ontological autonomy. Wilson strategy, by contrast, consists in accepting the first four premises about higher-level phenomena, denying alternatively the two principles. On the one hand, Strong metaphysical emergence, deriving from the satisfaction of the New Power Condition, denies the Physical Causal Closure admitting the emergence of new fundamental and distinctively efficacious higher-level properties. Weak emergence, on the other hand, depending on the fulfilment of the Proper Subset Condition, rejects Non-overdetermination thanks to the token-identity of higher-level powers with (a subset of) lower-level powers. Despite admitting this token-identity relation, however, the different power profiles of higher-level and lower-level features allows for the ontological and causal distinctness of special science entities (different entities exerting the same powers). In other words, the identity between powers involves an intimate relation of realisation between features, rendering what might seem over-determination something weak and unproblematic.

Eventually, the rejection of the Physical Causal Closure implies that Strong emergence is incompatible with physicalism, while the token-identity entailed by the Proper Subset Condition does not compromise the compatibility of physicalism and Weak metaphysical emergence.

1.5 *Conclusions*

The brief overview we made highlights a number of criteria for emergence which, although declined in different forms, recurrently appear in emergence literature. These criteria, as already said, are three. The first is *ontological irreducibility*, namely the impossibility of simply breaking down an emergent entity, property, or power into smaller, more fundamental elements without losing some of its essential features in the process. The second is *epistemological irreducibility*, namely the impossibility of exhaustively explaining, predicting, compressing, or deducing the nature of an emergent entity, property, or power starting from knowledge about its components alone. The third is *novelty*, which refers to different features, such as the presence, in emergent phenomena, of new, fundamental properties exerting new, fundamental powers, often ruled by new laws and conceptualized through new theoretical frameworks.

1.5.1 *Ontological irreducibility*

Ontological irreducibility, which opposes ontological emergentism to ontological reductionism, is an essential criterion in many models of emergence. As we saw, given the connection between emergence and the failure of generative atomistic physicalism (GAP), ontological irreducibility and decomposability is central in Humphreys' descriptions of emergence, and the same can be said of Gillett, for whom emergence appears when the Simple View of Aggregation, and consequently Fundamentalist physicalism, fails, namely when compound entities are composed by parts which contribute powers in virtue of their being part of a more complex structure, rather than in virtue of their nature alone. In Gillett, the Conditioned View of aggregation attributes determinativeness to relations and structure, and this kind of complex aggregation makes reduction impossible, allowing for ontological emergence instead. In Wilson, eventually, reductive physicalism is ruled out by the causal distinctness of emergent goings-on, which exhibit new fundamental powers, in the case of Strong emergence, or new causal profiles, in the case of Weak emergence.

Other authors too refer to irreducibility as a criterion for emergence[103] and this tendency to identify emergent entities with entities that *resist* to reduction is so common that the expression "the whole is more than the sum of its parts" has become the traditional formula

---

[103] See for instance Silberstein and McGeever 1999, Wimsatt 2000, Kim 2006. I will extensively talk about this in Chapter II.

for designating the meaning of emergentism. As we will see in the next chapter, however, mereological irreducibility is not enough to define emergence, given that, just like emergence, reduction too is not at all a clear, uncontroversial concept, and it is used in different ways depending on different contexts. As we will see in the very next paragraph, this circumstance is significant for epistemological reductionism as well.

1.5.2 *Epistemological irreducibility*

Epistemological irreducibility is a criterion we find in the first authors who wrote about emergent properties, namely John Stuart Mill and Charlie Dunbar Broad. As a matter of fact, Mill never used the term emergence, but he can be considered the father of emergentism because of his attention to what he defines "heteropathic effects".

In the sixth chapter of the third book of *A System of Logic*,[104] whose title is *Of the Composition of Causes*, Mill distinguishes between *homopathic* and *heteropathic* effects, reporting the essential difference, in nature, between those effects that are the sum of their causes, and those which are "heterogeneous to them".[105] For Mill, while mechanical laws are homopathic (in other words, linear or vectorial), chemical laws – and forces – are heteropathic, therefore their effect cannot be predicted before observing them. It is because chemical effects are heteropathic that Mill says that water properties are *deductively* unpredictable[106], and the same principle reappears in C.D. Broad, who talks about *transordinal laws*, rather than heteropathic laws, but still highlights that in order to discover the law at issue, "an actual instance"[107] of the higher-level object governed by the law is required.

The distinction between possible induction and impossible deduction of transordinal/heteropathic phenomena is central in Jaegwon Kim too, as we saw. With regard to emergent properties, he highlights the difference between *inductive predictability* and *theoretical predictability*[108], stating that even emergent properties can be predictable, but just inductively, which means after having collected empirical data and evidence.

[104] Mill 1843.
[105] Mill 1843: 431.
[106] See Mill's quotation in 1.2.2.
[107] Broad 1925: 79.
[108] Kim 1999: 8.

As for Humphreys and Gillett, they also formulate definitions of emergence in similar epistemological terms when they talk, respectively, of inferential, conceptual, and weak emergence.

A last remark is required, however. Several among the aforementioned formulations of epistemological criteria for emergence are, precisely, just criteria. Mill's and Broad's accounts of emergence, for instance, are metaphysical, but to identify emergent processes they choose to focus on epistemological marks. It is worth noticing, therefore, that the adoption of epistemological criteria does not necessarily undermine the metaphysical nature of emergent properties, reducing them to processes we simply cannot explain. The epistemological irreducibility criterion can (and should) be paired with other criteria, so that a more accurate comprehension of emergence can be obtained.

1.5.3 *Novelty*

Talking about emergence, novelty is often mentioned. In the mainstream view, genuine emergent properties are those properties which exhibit novel features not had by their parts. Novelty, however, can be defined in different ways.

At the beginning of the debate about emergence, metaphysical novelty (and, as a consequence, strong ontological emergence) was connected with nonlinearity[109], but in the last decades, complexity science highlighted that many complex systems exhibiting nonlinear behaviour are perfectly deterministic and do not involve any new fundamental force or property,[110] so it is not straightforward that nonlinearity implies emergence. Novelty is the central criterion for Humphreys' model of ontological emergence[111] as well, but for Humphreys, differently from British emergentists, an entity is novel "with respect to a domain D just in case it is not included in the closure of D under the closure criteria C that are appropriate for D".[112] We said that in this frame novelty is a relational property had by an entity that fails to be comprised under a certain domain which is defined by its closure under certain criteria. The definition of these criteria, therefore, are required to clarify the novelty at issue.

[109] We discuss about nonlinearity in the third chapter. See 3.2.
[110] About this see Wilson 2019: 206 et seq.
[111] See 1.2.1.
[112] Humphreys 2016: 29.

Eventually, novelty is significant in Jessica Wilson's work too, when she demands the fulfilment of the New Power Condition to have Strong Emergence. The notion of novelty Wilson exploits in her work is that of *fundamental* novelty, and the choice to connect emergence to fundamentality reflects a tendency which is shared by many authors.[113] This idea implies that the novelty emergent entities exhibit should be accepted as a "brute fact"[114] which requires no further justification. In this frame, the novelty possessed by emergent features (and powers) is *primitive* and this "ontological basicness" is in some sense absolute.

1.5.4 *Why further analysis is required*

The aforementioned criteria succeed in describing significant aspects of emergent phenomena, but additional clarifications appear to be necessary because, as a matter of fact, each of these criteria rests upon further theoretical backgrounds which should be analysed as well.

On the one hand, the first two criteria, ontological and epistemological irreducibility, describe emergence as something *resisting* to reduction. This is a negative definition, answering the question "what emergent entities are not" with the answer "emergent entities are entities that cannot be successfully reduced". To define emergence, instead, it is required a positive definition able to identify the features a phenomenon should have to be defined as emergent. Moreover, as we will see, this negative approach is nothing more than explaining an unclear concept through an equally puzzling notion. Just like emergence, as a matter of fact, reduction is not at all a clear, uncontroversial technical term, and it is used – and considered – in different ways depending on different contexts and theoretical frameworks. Throughout the last few decades, both reductive ontological physicalism and inter-theory reduction underwent several reformulations and adjustments, and there are still no univocal, empirically consistent models of them. Identifying emergence via failures of

[113] See for instance Cunningham 2001 (even if he uses the term "basic", rather than "fundamental"), O'Connor and Wong 2005, Barnes 2012, Bennett 2017.

[114] This point (and the expressione "brute fact") traces back to Samuel Alexander's *Space, Time and Deity*, in which he states that: "The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I would prefer to say in less harsh terms, to be accepted with the 'natural piety' of the investigator" (Alexander 1920: 46-47). The same expression is used in Broad as well: "[…] there are certain ultimate differences in the material world which must just be accepted as brute facts" (1925: 55).

reduction, therefore, seems a way to contaminate the former with the inaccuracies troubling the debate about the latter.

On the other hand, the third criterion, novelty, is defined as a not further analysable primitive in the case of fundamental novelty, or in opposition to physical acceptability or closure, in the case of nonlinearity, qualitative novelty and causal novelty. In the latter cases, emergence is described, again, in opposition to something else, namely the physical and its specifications. Adequate answers to the question *what is physical*, however, are far from being available, and the tendency to consider emergent phenomena as non-physical or non-natural seems moreover definitely anachronistic, given the extensive use of the concept in natural sciences.

CHAPTER II

Irreducibility

2.1 *Introduction*

In this chapter I focus on the first two criteria individuated in Chapter I, namely ontological and epistemological irreducibility, showing that both of them are rooted in further notions which are often accepted without being appropriately problematized. In the first case, quite obviously, the criterion identifies emergent entities with something resisting to ontological reduction in their being (i) *indecomposable in* and (ii) *causally irreducible to* their parts. In the second case, as far as epistemological reduction is concerned, emergents are (i) *non-deductible* and (ii) *unexplainable* starting from the knowledge of their parts. Ontological and epistemological irreducibility, therefore, rest quite obviously on the notions of ontological and epistemological (or *representational*) reduction, intended in the terms clearly stated by van Gulick in 2001:

> Between *what types of things* might the reduction relation hold? […]
>
> - a relation between real-world items — objects, events, or properties — which we might term Ontological Reduction (ONT-Reduction).
>
> or as
>
> - a relation between representational items — theories, concepts or models — which we can call Representational Reduction (REP-Reduction).[125]

[125] Van Gulick 2001: 2-3.

Reductionism, both ontological and epistemological, is supposed to be a method or a process broadly accepted in science, and many philosophers have considered it the keystone to advance and progress in knowledge. As we will see, however, there is not a unique, clear definition of reduction, therefore talking about reductionism in general is vague and misleading, and using it to define emergence is equally inaccurate. Let's now analyse the notions of ontological and epistemological reduction, so to frame the relationships between them and the idea of emergence.

2.2 *Reduction*

Talking about Humphreys' taxonomy, we highlighted that the notion of emergence is now widely used in science, and this fact should suggest that, firstly, emergence is not anymore[126] an empirical hypothesis ruled out by the progress of scientific knowledge, and, secondly, that the notion of emergence should not be obstinately opposed to reductionist views, which are undoubtedly equally essential to scientists. This idea may seem bizarre, because most philosophical literature[127] bears witness to the widespread habit to identify emergent entities with entities that resist to reduction, being it *ontological* reduction in the case of strong or ontological emergence, or *epistemological* reduction in the case of weak, inferential, or epistemic emergence (bur also in the case of ontological one, given that it might be said that ontological irreducibility implies epistemological irreducibility): in other words, it is almost as if one of the *conditio sine qua non* of emergence was the failure of reductionist decomposition or explanation.

David Chalmers, for instance, in defining strong, ontological emergence, focuses on the concept of *deducibility*, stating that strong emergence involves phenomena for which epistemic reduction fails:

> We can say that a high-level phenomenon is strongly emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain,

[126] The deeply influential survey provided by McLaughlin in 1992 outlines a narrative for which British Emergentism was justified by certain lacunae in the scientific knowledge, but when this knowledge became more complete, its ability to explain previously mysterious phenomena weakened the alleged usefulness of the notion of emergence. Nowadays, on the contrary, scientists themselves are those who use so diffusely the notion, so the framework seems significantly different.

[127] For a review, see Silberstein 2002, Hohwy & Kallestrup 2008, Bedau & Humphreys 2008, O'Connor & Wong 2015, Humphreys 2016.

> but truths concerning that phenomenon are not deducible even in principle from truths in the low-level domain.[128]

Jaegwon Kim's statements on the topic are similar, although his notion of reduction has changed over time from a Nagelian to a functionalist one:

> The distinction between properties that are emergent and those that are merely resultant is a central component of emergentism. As we have already seen, it is standard to characterize this distinction in terms of predictability and explainability.
>
> 3. *The unpredictability of emergent properties*: Emergent properties are not predictable from exhaustive information concerning their "basal conditions". In contrast, resultant properties are predictable from lower-level information.
>
> 4. *The unexplainability/irreducibility of emergent properties*: Emergent properties, unlike those that are merely resultant, are neither explainable nor reducible in terms of their basal conditions.[129]

Humphreys, in his already mentioned comprehensive work on emergence, writes that "emergence is usually taken to be incompatible with reduction in the sense that A's being irreducible to B is necessary condition for A to emerge from B"[130], and Michael Silberstein's opinion on the literature about the topic is similar as well. Summarising the traditional view about the relationship between emergence and reduction he states that: "'Emergentism', historically opposed to reductionism, is the 'ism' according to which both ontological and epistemological emergentism are more or less true, where ontological and epistemological emergence are just the negation of their reductive counterparts".[131]

Finally, it is important to note the use that Jessica Wilson makes of emergentism in a 2005 article dedicated to physicalism. In wondering whether physicalism can be based on the relationship of supervenience, which we will discuss later, Wilson sets a relevant requirement, which she calls the "criterion of adequate contrast". That is, an appropriate model of physicalism must be incompatible with its "best traditional rival"[132], i.e. emergentism.

---

[128] Chalmers 2006: 244.
[129] Kim 1999: 21.
[130] Humphreys 2016: 185.
[131] Silberstein 2002: 81.
[132] Wilson 2005: 430.

Although reductionism *per se* cannot be directly connected with physicalism,[133] in the emergence/reduction philosophical debate it seems the case: a mental property, for instance, seems emergent unless it can be exhaustively *reduced* to some neurophysiological processes. The issue, however, is more complicated than that.

As some philosophers of science and scientists have noticed[134], defining emergence as a reductionist failure is nothing more than explaining an ambiguous concept through an equally puzzling notion. Just like emergence, as a matter of fact, reduction is not at all a clear, uncontroversial technical term, and it is used and intended in different ways depending on different contexts. Patricia Churchland, for instance, says as follows:

> 'Reductionism' is a term of contention in academic circles. For some, it connotes a right-headed approach to any genuinely scientific field, an approach that seeks intertheoretic unity and real systematicity in the phenomena. It is an approach to be vigorously pursued and defended.
>
> For others, it connotes a wrong-headed approach that is narrow-minded and blind to the richness of the phenomena. It is a bullish instance of 'nothing-butery', insensitive to emergent complexity and higher-level organization. It is an approach to be resisted.[135]

Reductionism, therefore, can be both an epistemological tool able to provide intertheoretic unity and systematicity in the study and analysis of a phenomenon, and a metaphysical eliminativist thesis of "nothing-buttery" that for certain authors ignores the complexity of nature.

As for epistemological reductionism, an author who clarified the debate is Daniel Dennett. Dennett coined the well-known expression "greedy reductionism" to distinguish good and bad reductionism:

> We must distinguish reductionism, which is in general a good thing, from *greedy reductionism*, which is not. […] There is no reason to be compromising about what I call good reductionism. It is simply the commitment to non-question-begging science without any cheating by embracing mysteries or miracles at the outset. […]. But in their eagerness

[133] See Van Riel and Van Gulick 2018.
[134] Wimsatt 1997, Gillett 2016, Silberstein 2012.
[135] Churchland 1992: 18

for a bargain, in their zeal to explain too much too fast, scientists and philosophers often underestimate the complexities, trying to skip whole layers or levels of theory in their rush to *fasten everything securely and neatly to the foundation* [emphasis mine]. That is the sin of greedy reductionism, but notice that it is only when overzealousness leads to falsification of the phenomena that we should condemn it.[136]

In Dennett's opinion, greedy reductionists' bias is clear: only the most fundamental is scientifically unquestionable, therefore reductionists are committed to the idea that everything should be *directly* explained and "securely fasten" to that. The problem with this position is that it ignores the levels of complexity existing in nature, and this point is precisely the one highlighted by Richard Dawkins too, when he talks about "precipice" reductionism (apologize for the long quotation):

Reductionism has become a dirty word in certain circles. There's a kind of reductionism which is obviously silly and which no sensible person adopts, and that's what Dan Dennett calls "greedy reductionism". My own version of it is "precipice reductionism". If you take something like a computer: we know that everything a computer does is in principle explicable in terms of electrons moving along wires, or moving along semiconductor pathways. Nobody but a lunatic would attempt to explain what is going on in terms of electrons when you use *Microsoft Word*. To do so would be greedy reductionism. The equivalent of that would be to try to explain Shakespeare's poetry in terms of nerve impulses. You explain things in a hierarchy of levels. In the case of the computer, you explain the top-level software — something like Microsoft *Word* — in terms of software one level down, which would be procedures, subprograms, subroutines, and then you explain how they work in terms of another level down. We would go through the levels of machine codes, and we would then go down from machine codes to the level of semiconductor chips, and then you go down and explain them in terms of physics. This orderly, step-by-step way — what I call step-by-step reductionism, or hierarchical reductionism — is the proper way for science to proceed.

Reductionism is explanation. Everything must be explained reductionistically. But it must be explained *hierarchically and in step-by-step reductionism* [emphasis mine]. Greedy reductionism, or precipice reductionism, is to leap from the top of the

[136] Dennett 1995: 82.

> hierarchy down to the bottom of the hierarchy in one step. That you can't do; you won't explain anything to anybody's satisfaction".[137]

Dawkins highlighted that a good reductionism should take into consideration the complexity of the real and its hierarchical, multi-level structured organization, but "precipice" reductionism, as suggested by its name, skips levels instead, and directly connects different phenomena across the board, being eventually deceptive and unproductive.

Now, as far as ontological reductionism in concerned, an author who traces distinction between vulgar reductionism and a more aware scientific view entailing, on the one hand, some sort of reductionist heuristic, and respecting, on the other hand, real complexity is William Wimsatt, who defined the first trivial reductionism as "aggregativity" or "nothing-buttism".

When Wimsatt uses the term "aggregativity", he means a particular type of compositional relationship, somewhat similar to the kind found in cases where reduction works, but more extreme:

> Like interlevel reductions, aggregative relations are compositional. But aggregativity requires more. System properties that are aggregates of parts' properties represent degenerate cases where the organization of parts does not matter: they are invariant over organizational rearrangements. It is—roughly—a reduction without a mediating mechanism[138].

Aggregated systems are therefore systems in which the structural relations between the parts are not relevant, but what matters is only the presence of those parts that can be aggregated in several different ways without the integrity of the system being affected. Imagine a pile of sand. It is composed of smaller and more elementary parts, i.e. the grains of sand. It is not necessary, however, for certain grains to be at the base of the pile and other grains to be at the top. If the grains of sand were mixed, there would always be a pile of sand. Now imagine a cell instead. Its parts are organised in a complex structure and the relationships between them are vital for the cell to continue to exist function as a cell. A cell is not a simple aggregate of mitochondria, Golgi apparatuses, centrioles, membranes and the like that generally occupy a certain portion of space. It is a structured system and if the

[137] Dawkins in Brockman 1995: 77.
[138] Wimsatt & Sarkar 2006: 698.

parts do not stay in place and do not do what they are supposed to do, it becomes something else. Now, if a pile of sand is an aggregate, a cell is not. And if both can be analysed from a reductionist point of view, only the sand pile can be described by the concept of aggregativity. According to Wimsatt, therefore, reductionism should not be assimilated to the thesis that complex entities are aggregates of more elementary entities, because aggregates are unstructured systems independent of the context in which they are inserted, whereas real natural systems are almost always structured and context-related and require much more complex explanations than what can be provided by a thesis attributing aggregativity.

In Wimsatt's view,[139] the kind of reductionism that should be really taken into account is not aggregativity, but the standard scientific practice of searching for mechanistic, decompositional explanations. Aggregativity, instead, is the property had by those systems whose features atomistically depend upon their part's properties alone. Aggregative systems' properties are invariant under any decomposition, addition, or re-aggregation because their nature is not subjected to structural constraints. In other words, aggregates are *context and structure independent*. But genuine, non trivial reductionism, should not be merely identified with aggregativity and it does not seem such an extreme position. Genuine aggregates, moreover, are quite rare in nature, therefore aggregativity should not be taken as a universal model of reductionism.

To sum up, we saw that many authors underline that what reductionism is is a complex question, and that trivial forms of both epistemological and ontological reductionisms (greedy or precipice reductionism or aggregativity) should be kept separate from more sophisticated accounts of reduction, and the criticalities of the former should not discredit the legitimacy of the latter.

With regard to the relationship between reductionism and emergence, there are three theses I would like to support in this text. The first is that although there are several models of reductionism, none of them is really adequate to define emergentism by opposition. That is, it is not enough to say that a phenomenon is emergent if it cannot be reduced to something else, however refined the model of reduction chosen. The reason for this is twofold: on the one hand, even the most refined philosophical models of ontological and epistemological reduction do not seem to do what they promise, i.e. provide an exhaustive reduction that explains all the characteristics of high-level (reduced) phenomena starting from the most

[139] See Wimsatt 2000.

elementary (reducing) ones. On the other hand, if we consider more realistic and moderate reduction models, which do not promise a metaphysical reduction of the entities or a complete deductibility of the theories, the emergence does not seem incompatible with them, since these reductions leave non-reduced residues worthy of further consideration. The second thesis I would like to support refers to this: emergence and reduction are not opposite, but complementary, and recognizing the presence of reducible and emergent traits in high-level phenomena allows for the elaboration of a more complete description of reality than that provided by reductionism or emergentism taken individually.

The third thesis that I would like to put forward descends from the first two and concerns the reasonableness of pluralism. It is clear that, depending on the model of reduction chosen, the emergence opposite to it will take on different characteristics. We have already noticed this by examining Humphreys' and Gillett's theories, which interpret reductionism in different ways (Humphreys in terms of atomism and Gillett in terms of aggregativity and functionalism); this different interpretation leads ro two visions of emergence that cannot be entirely superimposed. Yet, this non-overlapping, as well as the different characteristics that the emergence will take on compared to the different forms of reductionism that we will consider, is a problem for the definition of the emergence only if we claim to provide one that is based on a fixed set of sufficient and necessary conditions - a definition like the one that we talked about in the introduction. If, on the other hand, we want to define emergence by using an open cluster of properties, it is not a problem to include different notions of irreducibility descending from different models of reducibility (i.e. reduction). Rather, this becomes further confirmation of the appropriateness of this theoretical move. There is no reason to decree that a single notion of irreducibility is appropriate, because the different reduction models found in the literature all have undeniable benefits. The different characterizations of irreducibility will therefore fall within the cluster and represent different ways in which an entity can meet the condition of irreducibility.

2.3 *Ontological reductionism*

Despite being usually connected to physicalism alone, ontological reductionism is, more generally, a monistic metaphysical thesis about the nature of the world affirming the existence of a single fundamental (kind of) substance that is causally efficacious – as opposed to the derivative entities composed by it whose causal powers are just “inherited”.

This substance, be it matter, energy, thought, God, or anything else, as well as being the only existing one, is also the source or the fundamental base of both the plurality of things populating reality and their properties and powers. When the fundamental substance is characterized in physical terms, this monistic metaphysics is called "materialism" or "physicalism"[140] and the history of the concept reveals different formulations introducing in the debate many other interesting questions, such as the problem of the unity of science, that of fundamentality, or that of the legitimacy of talking about levels of reality/being/organisation.

2.3.1 *Supervenience-based physicalism*

In Chapter I, physicalism was presented in Paul Humphreys' and Carl Gillett's terms. On the one hand, Humphreys describes a view of nature called Generative Atomism (GA) entailing, ontologically speaking, that "[…] all entities are either atoms or are composed of atoms".[141] Assumed GA, therefore, all existing entities should be decomposable into atomic entities.[142] In case of *physical* atoms, Humphreys talks about GAP (Generative Atomistic Physicalism), which is a model describing several physicalists views already present in literature. GAP entails that reality can be explained in generative/combinatorialist terms, and, in more complex circumstances, by means of supervenience: "[…] once the elements and structure of the lowest level are present, the objects, the structure, and the laws of the remaining levels are fixed by the individuals and the properties of the lower levels".[143] Gillet, on the other side, talks about Fundamentalist Physicalism, a view entailing a particularly strict notion of aggregativity, described by the so-called "Simple View of Aggregation". Fundamentalism, in Gillett terms', is "A global view that takes all entities in nature to be identical to, or compositionally explained by, the entities of microphysics, but further claims that only the entities of microphysics are determinative or exist".[144] Gillett's description of reductionism, however, does not focus on mereological composition alone, but also on

---

[140] For practical purposes, we will consider the terms as synonymous, even if they are not. See Stoljar 2017.
[141] Humphreys 2016: 12. The complete quotation is the following: "Generative atomism in this basic form has both a synthetic and an analytic component. The synthetic component says (1) that there is a collection of elementary entities from which all other legitimate objects in the domain are constructed, (2) there is a fixed set of rules that govern the construction process, and (3) as a consequence of (1) and (2), all entities are either atoms or are composed of atoms. The analytic component asserts that any non-atomic object can be uniquely decomposed into its atomic components using an explicitly formulated set of decomposing rules".
[142] See footnote above.
[143] *Ivi*: 3.
[144] Gillett 2016: 358.

causal realisation and on other kinds of compositional relationships, as we saw in § 1.3.4.1. About causal realisation, Gillett states that "given the nature of realisation relation, we can account for *all* the causal powers of individuals simply using the contributions of powers by the realizer property instances of these individuals".[145] In other words, in the mainstream view of causal realisation Gillett is analysing, only the realisers do the causal work and are causally efficacious, so only the realisers should be encompassed in our ontology.

This last view is similar to the definition of microphysicalism proposed in 1993 by Philip Pettit.[146] However, before outlining Pettit's account of physicalism, a premise is necessary. As rightfully pointed out by Tim Crane and David Hugh Mellor, physicalism has been adopted without a proper definition for a long time. A clarification, however, is necessary, because "The claim that everything is physical is not as clear as it seem".[147] Crane and Mellor notice that many authors consider physicalism as the modern version of materialism, but while materialism was a *metaphysical* doctrine about the very nature of physical stuff (it has to be "solid, inert, impenetrable and conserved, and to interact deterministically and only on contact"[148]), physicalism is more "subservient" and implies that "the empirical world […] contains just what a true complete physical science would say it contains".[149] The problem that arises at this point, however, concerns the identity of this science. Physics is obviously included in it, but it is less clear whether chemistry and molecular biology, for example, should also be included. In order to overcome this problem, physicalists usually suggest that the physical science in question must include physics and everything that can be reduced to it. Yet, this definition does not seem to be conclusive. On the one hand, it generates the problem of properly defining how one science can be reduced to another; on the other hand, it generates a further impasse in understanding what kind of physics should be the science that other disciplines ought to refer to: physics as it is now or some future physics[150]? If one considers current physics and posits that only the entities recognised by it exist, then physicalism is trivially false, since today's physics is undoubtedly incomplete and inaccurate, and basing physicalism on it would mean denying the existence of all that will be discovered in the future. The physics on which physicalism is based is therefore not the incomplete physics that is available to us today - if it were, then physicalism would be trivially false.

[145] Gillett in Corradini & O'Connor 2007: 204.
[146] Pettit 1993.
[147] Crane & Mellor 1990: 185.
[148] *Ivi*: 186.
[149] Ibidem.
[150] Crane & Mellor 1990: 188.

The alternative is therefore to base physicalism on the physics of the future: an ideal, complete physics that can explain everything. In this second case, however, three possible theoretical outcomes are generated, as Wilson points out, namely that physicalism is an indeterminate theory, as so is the physics of the future; that it is trivially true, since postulating that one day physics will explain *everything* seems like a convenient move, but completely abstract and vague; or, finally, that it will include entities that are normally excluded from it, such as the mind. If one assumes that one day physics will be able to explain everything, in fact, one cannot exclude the mind from this. In other words, it cannot be excluded that one day physics will admit the existence of a fundamental mental level. This last point, however, is unacceptable to Wilson, as it is to another philosopher who has dealt with these issues at length, David Papineau. Wilson and Papineau think that the physical, rather than being defined by present or future physics, must be characterised as being fundamentally non-mental, non-sentient and non-intentional. This vision is consistent with the idea highlighted by Crook and Gillett that physicalism is a modern version of materialism, where materialism was not a physical theory, but a mainly *metaphysical* theory that characterized reality in fundamentally non-dualistic terms and therefore implied that mental entities were not among the ontologically fundamental entities of reality.[151] This "fundamentally non-mental" characterisation of the physical is a controversial idea that has been widely debated and I will discuss the most relevant problems related to it in paragraph 3.2. For now, I would like to focus instead on a more classical view of physicalism, which is similar to Humphreys's view from which we started, and has been summarised by Philip Pettit.

Pettit believes it is possible to give a positive and non-trivial description of this thesis based on the following four assumptions. The first is a claim of realism: for the physicalist, there *really* are microphysical entities such as the ones assumed by physics, and this implies that physics is able to really describe reality. The second claim states that everything in the world is composed "without remainder"[152] by the aforementioned microphysical entities. Pettit does not specify which kind of composition is here involved (there are, rather, many acceptable kinds), but he clarifies that the composition at issue is conservative and "non-creative", by which he means that, given the same microphysical configuration, there will be no differences at all at the macro-level. These second point is outlined in the expression

[151] Ivi: 348.
[152] *Ivi*: 215.

"No macrophysical difference without a microphysical one":[153] a very common formula representing the meaning of supervenience. The third point about physicalism concerns physical regularities and laws. For Pettit, microphysical laws are primitive, and govern microphysical entities behaviours. In addition to them, however, no physicalists (expect for the eliminativists) would deny the existence of macro-level laws – chemical, biological, or psychological laws, for instance. Despite admitting their existence, though, the physicalist will not consider them as independent from the microphysical ones. This is the fourth and last claim about physicalism: macro-level laws "do not complement micro-level laws, taking up some degree of slack left by those laws, and […] they are not independent of micro-level laws: they do not have the potential to conflict with them and they do not serve to reinforce them". [154] Pettit's conclusion is therefore that "Not only is the empirical world microphysically constituted, the empirical world is also microphysically governed".[155]

As mentioned above, Pettit's doctrine of microphysicalism implies supervenience, namely the impossibility of *macro*physical differences without corresponding *micro*physical differences. This idea is common, and many authors exploit it to describe a physicalist worldview.[156] One of them is David Lewis who wrote: "If two possible worlds were exactly isomorphic in their patterns of coinstantiation of fundamental properties and relations, they would thereby be exactly alike *simpliciter*".[157] As well as Pettit, Lewis thinks that physics is the discipline bound to identify fundamental properties and relations, which are universally effective because they "[…] occur in the living and the dead parts of the world, and in the sentient and the insentient parts, and in the clever and the stupid parts".[158] Fundamental properties and relations, therefore, are physical; as a consequence, everything is either physical (fundamental), or supervenient on the physical (derivative). This view entails that by fixing the micro-level, every other level would be fixed. An often used metaphor to explain this point is theological; it has been originally formulated by Saul Kripke in *Naming*

[153] *Ivi*: 216.
[154] *Ivi*: 217.
[155] *Ivi*: 219.
[156] The problem of the relationship between the physical and the non-pyshical is, after all, one of the two big issues of reductionism. See van Gulick, 2001: 2 "[…] the notion of reduction is ambiguous along two principal dimensions: the types of items that are reductively linked and *the nature of the link involved* [emphasis mine]. Thus, to define a specific notion of reduction, we need to answer two questions:

- Question of the relata: Reduction is a relation, but what types of things does it link?
- Question of the link: In what way(s) must the items be linked to count as a reduction? [emphasis mine]".

[157] Lewis 1992: 51
[158] *Ivi*: 52.

and Necessity[159] and then adopted by many authors.[160] The point of the metaphor is that microphysical fundamental entities are the only ones God has to create to have the world She wants to have. By creating the microphysical, all the rest follows because, given that, all the rest necessarily obtains. In Frank Jackson terms: "Any world which is a minimal physical duplicate of our world is a duplicate *simpliciter* of our world"[161].

To better understand these last points, we could think about how a television works, that is how the distribution of different coloured pixels on a screen generates images and shapes that the human eye (if placed at a proper distance) perceives as having gestaltic properties and a semantic value. If you turn on the television you'll see people, animals or objects, and you can't help but see those people, animals and objects unless you get close enough to the screen to see the coloured pixels that make up those figures. This example recalls the one provided by David Lewis in 1992,[162] about a grid of pixels and the corresponding images. This type of example is interesting because it is very intuitive. If reality were comparable to a television, the pixels would correspond to the fundamental properties and the images to the macroscopic, not fundamental ones. Now, although anyone would agree that when I see an object on TV that object is somehow there, it is still nothing more than the effect of the corresponding pixels on the screen. Without the pixels, there would be no image, while given the pixels, there can only be an image (and especially *that* image). The meaning of the supervenience, in this case, becomes obvious: there are no macroscopic changes (read: there can't be a different image on the screen) without microscopic changes (read: without the pixel distribution changing).

Considering what has been said so far, i.e. the very close correlation between physicalism and supervenience, the first criterion that is commonly used to define emergence, i.e. ontological irreducibility, could be seen as coinciding with the failure of supervenience. In fact, we have seen that for Humphreys there is emergence when Generative Atomism fails and this means, in fact, that emergence occurs if there is no appropriate supervenience

---

[159] See Kripke 1972: 153 et seq.

[160] See Schaffer 2004: 100 "[…] 'all God had to do' was to create the primarily real [the fundamental]"; Shoemaker 2007: 33 "If God wants to create a world like ours, there is nothing he need do beyond creating the sorts of micro-entities there are in our world, giving them the properties they have in our world, distributing them as they are distributed in our world, and laying down the laws that in our world govern the interaction of these entities"; Barnes 2012: 826 "The fundamental entities are all and only those entities which God needs to create in order to make the world how it is"; Wilson 2014: 540 "The fundamental is, well, fundamental: entities in a fundamental base play a role analogous to axioms in a theory – they are basic, they are 'all God had to do, or create'".

[161] Jackson 1998: 13. See also Loewer 2001: 38 et seq.

[162] Lewis 1992: 53.

relationship between the basis of supervenience and the higher level supervening objects that derive from it. There is emergence, in short, if the basic microphysical configuration is not sufficient to determine the nature of the resulting phenomena. However, at this point, an important consideration should be made. When we talk about relations of determination or dependence it is absolutely essential to specify their modal force. In other words, a dependency relationship may be contingent or necessary.

Let us give an example by taking causal relationships. The effect of a given cause can be contingent, in the sense that it may occur, or necessary, in the sense that it cannot fail to occur.[163] In the first case, given cause c, it is *possible* there may be effect f; in the second case, given c, it is *necessary* that f is produced (i.e.: it is not possible that f is not produced). In turn, moreover, a necessary effect may be necessary in different ways: its necessity may be nomological or metaphysical.[164] In the first case, where the effect is nomologically necessary, it occurs because, given the natural laws in place (and the right context conditions), it cannot fail to occur: its occurrence depends on its cause, but also on how the world is nomologically constituted. This implies that in a nomologically different world, i.e. one endowed with different laws of nature, that same effect may not occur.

When an effect is metaphysically necessary, then it cannot fail to occur regardless of any external conditions. No matter what laws of nature are valid within the world in question or what contextual conditions are given: if the effect of a given cause is metaphysically necessary, then, given that cause, that effect will be produced by necessity. It remains to be understood why the nomological structure of the world and environmental conditions are insignificant for metaphysically necessary effects. The reason for this is that when a relationship of dependence is of a metaphysical order, it is rooted exclusively on the intrinsic individual properties and powers of the entities involved in the relationship; any other factor, such as the relationships that the entities in question may have with the surrounding reality, is not relevant. If A and B are linked by a relationship of metaphysical dependence that implies B given A, then if A B will follow, regardless of any context condition, which means in every possible world.

Going back to the starting point, i.e. supervenience, it will now be clear that the modal force of this relationship is relevant to the discussion on the relationship between supervenience and emergence. We have seen that a relationship of dependence between two

[163] For more on this see Wolf 1957 (necessary contingent effects).
[164] Necessity may also be logical, but I will leave this point out because it is irrelevant to our discussion.

entities in which the former needs the latter can be valid by virtue of (at least) two different forms of necessity: nomological or metaphysical. We have also seen that the first form of necessity depends on the laws of nature of the world in which the entities involved exist, while the second is rooted in their intrinsic nature. Returning now to emergent phenomena, it cannot be said that they are not supervenient on their emergence basis, because if the supervenience is declined in terms of nomological necessity, then emergent phenomena can be said to be supervenient. These phenomena, in other words, emerge from their emergence basis and exhibit their typical properties according to the laws of nature that govern this world. If natural laws were to change or if we found ourselves in a possible world characterised by different laws of nature, emergent phenomena might not occur because they would depend not only on the individual properties of their constituent parts, as we shall see shortly, but also on their relational structures, which in turn depend on the natural laws in force and on the context (i.e. the natural world and its configuration as an organic whole).

A relationship of metaphysically necessary dependence, on the contrary, remains valid in all possible worlds because it depends exclusively on the intrinsic and essential properties of the parties involved and these properties, in turn, as essential, remain stable in all possible worlds. If supervenience is therefore understood in terms of nomological necessity, emergent phenomena can be defined as supervenient. If the supervenience relationship is understood in terms of metaphysical necessity, on the contrary, it will be clear that emergent phenomena cannot be considered as supervenient. It happens, however, that the supervenience relations to which physicalism refers are historically *metaphysical* supervenience relations, and it is for this reason that physicalism seems incompatible with emergentism.

It should be noted, however, that an author like Jessica Wilson[165] argues that this distinction between metaphysical necessity and nomological necessity is not always valid because it the nature of an entity may depend exactly on the laws of nature to which it is subject and therefore, in such a case, there is no distinction between a necessity that depends on the essential nature of the entity and a necessity that depends on the laws of nature governing that entity. Wilson believes that there are very good reasons to think that many individual physical properties actually depend on the laws of nature and therefore she does not worry too much about this distinction between metaphysical and nomological necessity in the definition of supervene on the one hand and emergence on the other. For Wilson, if the individual properties of phenomena depend *essentially* on the nomological structure of

[165] See Wilson 2005.

reality, then emergent phenomena will supervene on their most fundamental configurations with metaphysical and not only nomological necessity. This is an interesting position, with which I agree, and it is another point in favour of the idea that the contrast between supervenience and emergence is not useful either to exhaustively clarify the nature of emergence nor to define physicalism and reductionism.

Now, this first perspective designates *physicalism as the conjunction between atomism and supervenience* and describes *emergence as a supervenience failure*. However, this model may seem unsatisfactory because of two reasons. The first is that this is a negative definition, which focuses on what emergent properties *are not*, leaving unsolved the problem of what they *are*. The second is that supervenience is an underdetermined relationship.[166] What supervenience entails is that, given a certain configuration of low-level properties *P*, another configuration of high-level properties *M* will follow. Nonetheless, supervenience does not clarify *in virtue of what* that happens; it just requires correlation and covariance between two sets of properties, but it could not help in explaining *why* this covariance obtains, being supervenience metaphysically neutral. As McLaughlin and Bennett notice about abstract correlations of this kind, they "are themselves not explanatory (rather, their holding calls for explanation)".[167] Supervenience describes a dependence relation between lower and higher-level properties, but this determination could be nomological (and, therefore, contingent), holding in our world by virtue of its laws of nature, or metaphysical (and, therefore, necessary), holding in all possible worlds by virtue of the essential nature of the involved properties. Without any further determination of the relation at issue, the covariance between the *M*-properties and the *P*-properties could even be casual, or managed by a supernatural entity like God. Given this ontological vagueness, and despite the many supervenience-based formulations of physicalism present in the literature, supervenience alone is not generally accepted as sufficient for physicalism,[168] therefore the first criterion for emergence should not be liquidated as a criterion just based on supervenience failure.

There is, however, another formulation of physicalism that tries to determine in more exhaustive terms the nature of the relationship between the physical and what is supposed to supervene on it. This formulation exploits the notion of realisation, sometimes called implementation. In this case, higher-level properties supervene on lower-level properties

[166] See Wilson, 1999.
[167] McLaughlin, Bennett: §3.7. See also See Wilson 2005: 433 "Supervenience is a relation aimed at capturing the dependence of one family of properties on another by means of correlations alone, rather than by spelling out the precise nature of this dependence".
[168] About this, Wilson 2005 provides a good review.

because of some realizers which render *constitutive* the relationship between the supervenience base and the supervenient phenomena. Realization, therefore, is supposed to clarify the metaphysical weakness exhibited by the notion of supervenience, and to explain the reason of the covariance correlated to it.

2.3.2 *Realisation-based physicalism*

In the first chapter, we saw Humphreys' and Gillett's accounts of reductionism and realisation. In literature, however, other accounts are present as well, and the realisation relationship is characterised in different ways.

Generally speaking, realization is supposed to be a dependence relation obtaining between higher and lower-level properties or states by metaphysical necessity. As shown by the literature, however, there are different relations which seem to fill this schema, and these relations can be defined "realization relations".[169] Despite their differences, all of them exhibit four common features. First, they are *asymmetric*: if an entity *A* is realized by an entity *B*, then *B* is not realized by *A*. Second, they are *irreflexive*: *A* cannot realize itself. Third, they are *transitive*: if *A* is realized by *B*, and *B* is realized by *C*, then A is realized by *C*.[170] Fourth, they are *synchronic*: *A* is synchronically realized by *B*, and this is supposed to rule out a possible identification between realization and causation, which is usually took as a diachronic relation.

Beyond these formal requirements, realization is characterised as a relation obtaining between "first-order" and "second-order" properties, *via* some kind of realizers. The distinction between the former and the latter has been well described by Shoemaker:

> […] A realized property is said to be a second-order property, and its realizers are said to be first-order properties. Since the properties that realize a property may in turn be realized by other properties, it might be better to say that the realized property is a higher-order property and its realizers are, relative to it, lower-order properties.[171]

[169] This is the suggestion of Umut Baysan. See Baysan 2015.
[170] It is worth noticing that this is not uncontroversial. Paul Humphreys (2016: 17), for instance, states the opposite: "I am a functional part of the economy of Charlottesville, and my liver is a functional part of me, but my liver is not (yet) a functional part of the Charlottesville economy".
[171] Shoemaker 2007: 11.

In other words, a higher-level property, such as a special science one, is realized by a first-order property, namely a more fundamental one (such as a physical one), *via* certain parts of the system in which the latter is instantiated that perform the role by which the second-order property is individuated. This implies that having a second-order property *consists in* having a first-order property playing the role at issue.

A clear example of this kind of relational structure is *functionalism*, which, historically, walked hand in hand with the problem of the mind and has a history that is strictly connected with the history of the notion of realisation.[172] In the functionalist framework, mental properties are second-order properties, while neural properties are lower, more fundamental ones. Kim's functionalism, for instance, implies that mental properties are realized by neural properties, but to do that appropriate realizers in the system in which neural properties are instantiated should be identified. This system is, patently, the brain (or the nervous system), therefore mental properties' realizers will be those parts of the brain which perform the functions, or causal roles, individuating mental properties. A functionalist like Kim, therefore, first identifies second-order properties with certain associated functional or causal roles, and then reduces them to the lower-level realisers of these functional roles. The process has been clearly described by Jaegwon Kim in the following three-steps reduction, where the second-order property *E* is reduced to the lower-level domain B through its "functionalisation":

> Step 1: *E* must be *functionalized* – that is, *E* must be construed, or reconstrued, as a property defined by its causal/nomic relations to other properties, specifically properties in the reduction base **B**. […]
>
> Step 2: Find realizers of *E* in **B**. If the reduction, or reductive explanation, of a particular instance of *E* in a given system is wanted, find the particular realizing property *P* in virtue of which *E* is instantiated on this occasion in this system; similarly, for classes of systems belonging to the same species or structure types. […]
>
> Step 3: Find a theory (at the level of **B**) that explains how realizers of *E* perform the causal task that is constitutive of *E* (i.e., the causal role specified in Step 1). Such a theory may also explain other significant causal/nomic relations in which *E* plays a role.[173]

[172] Shoemaker 2007: 2 et seq.
[173] Kim 1999: 10-11

This three-steps functionalisation – to which Kim add a further step saying that *E* should be identified with its realizer on some occasion – fields an interesting but questionable strategy. First of all, it is important not to forget the real purpose of any reduction, which is to reduce (and explain[174]) an unclear phenomenon – a property, in this case – to (and through) its constituent parts, which are usually less metaphysically controversial and better known. In this case, though, the original phenomenon *E* – a mental state – is immediately reshaped and modelled taking into account its "causal-nomic relations" with other phenomena alone, and ignoring, conversely, its other properties. It could be reasonable, therefore, questioning the legitimacy of Step 1. What should be at least justified is the reason why the causal/functional role of *E* is sufficient to provide an exhaustive definition of *E*, and why all *E*'s other features can be ignored. It is worth noticing that in general functionalism works for properties whose natures are exhausted by their causal roles, that is, for properties that really are functional. In those cases, there is no reshaping, but just a process during which the causal role that defines the properties at issue is rendered explicit. In the case illustrated by Kim, by contrast, the functionalised property cannot be reasonably considered exhaustively defined by its causal role. In fact, a further question could be why causal or functional properties should be considered the most metaphysically prominent properties in the definition of the identity of the phenomenon Kim analyses. As already said, functional reduction is often applied to complex, mental phenomena, but this is an inappropriate strategy given the peculiarity of these phenomena. Focusing on the functional character of mental states is a deliberate a-priori choice corresponding to the exclusion of other characters which nonetheless seem as essential to mental phenomena as their functional features – for instance, the qualitative/phenomenological features of mental states, as well as their meanings, i.e. their semantic character. It is worth noting, in this respect, that Kim himself admits this difficulty in his book *Physicalism or something near enough*, when he considers the problems left open by physicalism: "Are mental properties physically reducible? Yes and no: intentional/cognitive properties are reducible, but qualitative properties of consciousness, or 'qualia,' are not"[175]. In Kim's opinion, however, this last point is not a huge problem. Physicalism can account for many properties, but what cannot save are "[…] intrinsic qualities—the fact that yellow looks like this, that ammonia smells like that, and so on. But, I say, this isn't losing much".[176]

[174] Baysan clearly highlights that realisation is an explanatory relation. See Baysan 2015.
[175] Kim 2005: 174.
[176] Ibidem.

To sum up, the first problem of functionalism is its lack of loyalty towards the original property *E*. Properties are *identified* with their having some functional roles, but this identification is anything but straightforward and seems to replace the original phenomenon with a surrogate. As a consequence, if the eventually reduced entity is something different from the entity we originally wanted to reduce, becoming what can be called *E**, the reduction of *E** to the realizer property *P* may even be successful, but it will reduce (and explain) *E**, rather than *E*, being therefore useless to the original purpose.

The identification of a property with its causal powers is essential to functional reduction, because if we proceed to Step 2 and 3, we can see that the identification of the original property *E* to the lower-level realizer *P* in B passes through the identification of *E*'s causal roles with *P*'s causal roles. Nonetheless, it seems a huge simplification to identify *E* with its causal powers, as *E* can have many other features contributing to its identity. Therefore, it seems that Kim's statement "[…] the *E*-instance and the *Q*-instance [*Q* is the instantiation of *P*] have identical causal properties, and this exerts powerful pressure to identify them"[177] is controversial, metaphysically speaking, because having the same causal properties does not imply being the same entity.

A related problem with functionalism is that in Kim's terms functional reduction is eliminative. Kim takes into account three possible interpretations of functional reduction, which are the following. The first one is conservative: "*First*, one may choose to defend *E* as a legitimate higher-level property irreducible to its realizers, the Q's";[178] the second one merely entails nomological necessity in the reduction of *E* to instantiations of *P* (i.e. *Qs*): "*Second*, one may choose to identify *E* with the disjunction of its realizers, Q1 v Q2 v… Notice, though, that this identity is not necessary – it does not hold in every possible world – since whether or not a property realizes *E* depends on the laws that prevail at a given world";[179] finally, the third one is instead eliminative: "*Third*, we may give up *E* as a genuine property and only recognize the expression "E" or the concept E".[180] Nonetheless, Kim considers the first two versions of functionalism unsustainable, and states that real scientific reductions are eliminative because higher-level properties are *scientifically irrelevant*.

The third and final problem with this kind or realization is connected with this last point: functionalism, in being unable to explain an autonomous mental causality, eliminates it in

177 Kim, 1999: 16.
178 *Ivi*: 16.
179 Ibidem.
180 *Ivi*: 17.

order to avoid the so-called "exclusion problem". If the relevant causal roles of the second-order properties are performed by the lower-level realizers in B, then no real causal powers can be attributed to the second-order properties, otherwise overdetermination would follow. This is a huge problem for realization, however, for it should be explanatory, and, therefore, should preserve the causal autonomy of second-order properties, rather than denying it. Among others[181], Shoemaker is particularly critical on this point, which he considers inacceptable[182]: "We need an account of property-realization that assigns the relevant causal role to the realized property itself, while acknowledging that it is in virtue of causal roles played by its realizers that it is able to play this causal role".[183]

Shoemaker, therefore, provides an account of physical realization that avoids the exclusion problem keeping distinct the second and the first-order properties. In his view, properties are individuated by their causal profiles, which consist of two kinds of causal features:

> […] forward-looking causal features, having to do with how the instantiation of the property contributes to producing various sorts of effects (and contributes to bestowing causal powers on its possessors), and backward looking-causal features, having to do with what sorts of states of affairs can cause the instantiation of the property.[184]

Now, both second and first-order properties have particular, individuating causal profiles, by which it is entailed that they possess different sets of forward-looking and backward-looking causal features. Realization obtains when, between these different sets of properties, holds a certain relationship defined as "subset account" of property realization. This relationship is the following. Given a second-order property *P* and a first-order property *Q*, *P* is realized by *Q* when (i) the set of *P*'s forward-looking features *is a subset* of the set of *Q*'s forward-looking features; (ii) the set of *P*'s backward-looking features *has as a subset* the set of *Q*'s backward-looking features.

The subset account implies that the causal powers of the second-order property *P* are not identical to the causal powers of the first-order property *Q*, being the former a subset of the latter, and this point implies that instances of the second-order property cannot be identical to instances of the first-order one. Therefore, the subset account, in Shoemaker's view,

[181] See Clapp 2001, Watkins 2002, Wilson 2019.
[182] Shoemaker 2007.
[183] *Ivi*: 5.
[184] *Ivi*: 12.

avoids the elimination of second-order properties, or at least the consequence of having epiphenomenal second-order properties without autonomous causal powers.

Shoemaker, however, highlights a further significant point. A good account of realization requires that the instantiation of the realizer of the second-order property *Q* produces the instantiation of *Q by necessity*. In several cases, however, finding a property realizer is not sufficient for the realization of the second-order property because of the relevance of other additional conditions. For instance, assumed that the first-order property "having C-fiber stimulations" realizes the second-order property "having pain", being these stimulations in a brain and not in a Petri dish is a necessary condition for the obtaining of the second-order property. In other words, it is common that a first-order property instantiation *alone* is not enough for the instantiation of a second-order property. What is rather required is the presence of a more complex microphysical state of affairs involving, at the same time, different property instantiations. Shoemaker distinguishes, therefore, between "property realization", where the realizer is a property instantiation, and a more fundamental sort of realization called "microphysical realization", where the realizer is a more complex microphysical *state of affairs*. Shoemaker defines these states of affairs as "the instantiation of micro-properties in micro-entities".[185] These instantiations determine in a constitutive way all the macrophysical facts about the world, namely both the instantiation of macrophysical properties and the existence of the macrophysical entities in which these properties are instantiated, and the reason for this is significant: "[…] plainly the instantiation of a property entails the existence of something in which it is instantiated".[186]

However, in admitting microphysical states corresponding to microproperties instantiated in micro-entities, the realization implies again, albeit in a theoretically more refined form, the metaphysical atomistic and fundamentalist picture found in the position criticized by Humphreys, Generative Atomism, and the one criticized by Gillett, Fundamentalist Physicism.

2.3.3 Conclusions

At this point I have to make a consideration that I consider relevant here: physicalism is clearly a "scientific" metaphysics, i.e. a metaphysics that accepts as existing only what is

[185] *Ivi*: 32.
[186] *Ivi*: 35-36.

admitted by the best scientific theory we have about a certain ontological domain.[187] However, the two most commonly accepted assumptions of physicalism - that there is indeed a fundamental level of reality and that there are ultimate unchangeable, fundamental entities constituting and governing all reality through their powers - are assumptions that physicists themselves find it difficult to commit to given the essentially speculative nature of these arguments. What physicists do, however, (an operation that, I believe, holds some metaphysical elements) is to define the objects of subatomic physics - those which at first glance would seem to be the most fundamental - as "dynamic objects". Such objects correspond to entities characterized by their relationships and dynamics, rather than by some persistent substantiality in time and space, and on this point it may be interesting to recall what Erwin Schrödinger wrote in the 1950s, in his book *Science and Humanism*, in which he tried to provide a popular description of the subatomic world. Schrödinger notes that if one observes a subatomic particle "here and now" and observes a similar one shortly afterwards in a very close place, not only is it unreasonable to think that those two particles are the same particle. Indeed, even wondering whether they are or not makes no sense in the world of subatomic objects, because they should not be considered permanent entities that can be somewhere, but rather simple *instantaneous events*.[188]

Let us underline this: not entities, but *events*, i.e. not a substance that persists and can therefore be individuated over and over again, but something that simply *happens* in instantaneous form without being able to repeat itself equal to itself because there is no substance and no "self" that persists between one occurrence and another. What physics, and specifically quantum electrodynamics, can tell us about the nature and dynamics of the subatomic world therefore indicates that microphysical reality consists of a continuous swarming of events that give themselves in dynamic and instantaneous form without permanent states or identities.[189] Starting from these considerations, I would venture to say that rather than physicalism, this scenario seems to support a widespread emergentism even at the level of physics, since the macroworld has somehow emerged from a microphysical reality made up of instantaneous, dynamic and non-permanent events. What emerged is a materiality characterized by persistence, impenetrability and stability - properties that do not belong to the subatomic domain and that are ontologically *new* with respect to it. But as I

[187] See Andina 2013: 30.
[188] Schrödinger 1952. On the debate on the non-permanence of subatomic entities the literature is very wide. See, in addition to Schrödinger, Cassirer 1956, Ladyman & Ross 2006, Lewis 2016.
[189] On this, see Feynman's book dedicated to introducing quantum electrodynamics: *QED. The strange theory of light and matter* (Feynman 1992)

said, this is only a ventured hypothesis (for now), because the issue is very complex. However, I think we can conclude that if nineteenth-century metaphysical materialism could enjoy a strong theoretical coherence with classical mechanics (and could be considered to all intents and purposes a "scientific" metaphysics corroborated by science), modern physicalism does not find an unconditional support in the evidence of quantum mechanics and contemporary physics, whose interpretation on a metaphysical level is not immediate and does not exclude any vision of reality that does not respond to the dictates of physicalists and reductionists.[190]

There is also a further difficulty that afflicts physicalism, in addition to the problems of atomism. If functionalism was only able to effectively reduce properties that can be exhaustively defined in functional terms, Shoemaker's realization is only effective in the case of properties that can be exhaustively identified by their individual causal profiles. After all, it is no coincidence that Wilson, in his metaphysics, resorts to realization: she declares from the outset that she is interested in the emergence of powers, because in her metaphysics the nature of properties and entities follows from the identification of the nature of powers. However, besides the functional properties and those that are actually identifiable from their individual causal profiles, there are other properties that remain excluded from this scenario.

Much like functionalism, in short, realization does not really seem adequate to provide an exhaustive explanation and reduction of complex and not immediately causal properties such as phenomenal, intentional, spatial, or qualitative ones. The problem does not only concern the mental domain, the traditional *bête noire* of philosophy, but all those processes and natural phenomena that cannot be defined in terms of individual causal functions or powers. It therefore seems that alongside these functional or causal properties, which undeniably exist, there are others, and if the analysed types of ontological reduction can effectively account for the former, something else is needed to account for the latter. This is why I do not think that emergence and reduction can be defined by opposition: they are not really incompatible or opposed. They simply describe different aspects of reality, and totally abandoning reduction in favour of emergence is as short-sighted and unreasonable as rejecting emergence in the hope that reduction can do all the work necessary to understand every aspect of reality.

[190] On this see, again, Lewis 2006. In this text, titled *Quantum Ontology*, Lewis provides a review of possible metaphysical interpretations of subatomic reality, showing that there are many different ones and that it is impossible, at least for now, to "decide" which one is the right one.

In this framework, it may be appropriate to develop a more liberal position that admits both a mereological and functionalist reductionism and emergence, which may be able to account for those properties that escape physicalist reduction. This "extended" vision, for which reality can be explained in some ways by reductionism and in others by emergence, seems more promising than simple micro-reductionism or realization, because a formulation so conceived could convey the idea that reality, while showing a compositional structure consistent with the premises of reductionism, also shows a certain discontinuity and novelty in the case of those properties that are not reducible to functional or directly causal ones.

The forms of reductionism that we have analysed so far characterised the natural world as a fundamentally continuous realm marked by "ontological unity" at the level of constitution and mereology, as well as at the nomological level (recall Pettit's assertion that the empirical world is not only *constituted*, but also *governed* by the microphysical domain).[191] Instead, an alternative vision could be to accept on the one hand the limits of atomistic and functionalist explanations (without denying their usefulness when they work), while also accepting the reality of those *differences* exhibited by the natural world at their various levels of organisation and complexity. Therefore, if it is true that (i) everything that is physical is *made* up of physical parts; (ii) that the individual functional and causal properties can be *explained* and *reduced* by resorting to their realizers and that (iii) the regularities of the microphysical world, conceptualized by their laws, are not affected by the regularities that can be observed at the most complex macroscopic levels, it seems that there is still a part of reality that escapes reduction and the physicalist model. What to do with this "irreducible" portion of reality? Recourse to emergentism could be a solution, and analysing this position as only partially opposed to reductionism seems to be even more reasonable.

At this point, and assuming that ontological reduction and emergence apply to different aspects of reality, I believe it makes sense to attribute equal ontological dignity to microphysical and macrophysical entities, properties and powers. The latter have a nature, characteristics and effectiveness that cannot be superimposed on those of the microphysical entities and therefore must be analysed in terms of emergence, as well as reduction. This extended vision, which aims to reconcile reductionism and emergence, will be better outlined in the last chapter. What remains to be done, however, is to understand whether this conciliation project is coherent with the epistemological reduction models in the literature. The next paragraphs of this chapter will therefore be devoted to understanding whether the

[191] Hüttemann 2004: 10.

epistemological reduction models developed in the literature are really capable of reducing our scientific theories to each other, providing a single knowledge of reality that rests and develops exhaustively (i.e. without residues) on the knowledge of the microphysical world.

2.4 *Epistemological reductionism*

As Manfred Stöckler pointed out, while the words "reduce" and "reduction" have been widely used in the past, "reductionism" is a word that originated in the second half of the twentieth century.[192] Its first uses, which date back to Ernest Nagel,[193] Willard Van Orman Quine[194] and John E. Hare,[195] suggest a type of reduction that involves systems of propositions and logical constructs rather than portions of reality.[196] The history of science has therefore offered models of reduction that involved cases of what would later be called "inter-theoretical"[197] reduction and these cases have been at the centre of many philosophical debates both for their methodological and epistemological value and for the ontological implications that have often been attributed to reductionism.

The simple statement we should assume as a starting-point is that, trivially, an inter-theoric reduction is a reduction holding between symbolic systems, i.e. theories, laws, statements, or scientific truths. In this sense, when we say that an emergent entity is *epistemologically irreducible*, we are stating that theories, laws, statements, or scientific truths concerning that entity cannot be translated in and explained by more fundamental theories, laws, statements, and truths. Along this basic idea, however, the debate is extremely complex, given that classical models of inter-theoric reduction have required more than that, such as the presence of bridging correlations connecting concepts of the old theory to concepts of the new one, the different degrees of fundamentality or the equivalence in observational prediction of the theories at issue, the possibility of a *complete* deduction and explanation of the old theory by the new one, and so on and so forth. Nonetheless, as we will see, there are cases in which bridging correlations cannot be found,

---

[192] Stöckler 1991: 72.
[193] Nagel 1949.
[194] Quine 1953.
[195] Lepre 1952.
[196] Si veda Stöckler 1991.
[197] For example, reductions involving the following pairs of theories / disciplines: thermodynamics and statistical mechanics (see Nagel 1961, Sklar 1993, Callender 1999), classical mechanics and special relativity (see Bohr's discussion of complementarity in Bokulich 2014, see also Nickles 1973), chemistry and quantum physics (see Scerri 2008, Hettema 2017).

cases in which it is not clear if the reduced theory is more or less fundamental than the reducing one, cases in which the theory reduction is cumbersome and merely partial, and no cases at all in which different theories appropriately "cover" and explain the very same range of phenomena. Sometimes[198] it is not even clear in which direction the reduction goes! All these circumstances have produced worries to scientists and philosophers of science, but nonetheless, outside science, epistemological reducibility is often taken for granted, despite being seldom fully considered. Inter-theoric reduction, as we said, is a semantic model of reduction, which means that reduction, here, must be viewed as a relation holding between epistemic representations of reality, rather than between ontological domains. In such a framework, "reduction" means "explanation",[199] and to illustrate this simple point, I will provide a brief description of some of these models, starting from the classical Nagelian one, which, although some ontological suggestions,[200] focuses on the reduction

[198] See Nickles 1973: 182 "Does classical mechanics (CM) reduce to the special theory of relativity (STR), or, on the contrary, does STR reduce to CM? Does Bohr's early quantum theory of the hydrogen atom reduce to classical theory of the atom in the mathematical limit of high quantum numbers, or, on the contrary, should we say that the classical theory reduces to (is reduced by) Bohr's theory? Surprisingly, the answer is not immediately clear". Here Nickles focuses on the difference between philosophers' and scientists' notion of reduction. While philosophers focus on a "domain-combining" reduction, in which the less general theory is reduced to the more general with a consequent unification of two previously different domains (an example of this is the reduction of physical optic to electromagnetic theory), scientists, by contrast, focus on a "domain-preserving" reduction, in which one of the theories involved reveals itself as a special case of the other, holding in a special limit. This is the case of CM and STR, in which we can say that STR reduces to CM in the limit of low velocities (with $v \to 0$), even if CM seems less fundamental of STR. It is worth noting, however, that Lawrence Sklar places the issue in the opposite way: "If we first restrict our attention to sentences framable in purely kinematic concepts, and then further restrict our attention to the subset of sentences in this class dealing with sufficiently low velocities, we find for every sentence in this subset derivable from the relativistic theory there is a sentence approximative to it derivable from the Newtonian theory. It is only in this extremely weak sense that Newtonian mechanics is reducible to special relativity" (Sklar 1967: 116). On this, see also Chibbaro, Rondoni & Vulpiani, 2014: 30 et seq.

[199] Nagel 1961: 338 "Reduction, in the sense in which the word is here employed, is the explanation of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain". See also Nagel 1970:

[200] As noted by some philosophers such as van Riel and Van Gulick (2018), some passages of Nagel's works provide ontological indications suggesting connections between theories and ontologies. See Nagel 1961: 339-40 "Difficulties are frequently experienced in comprehending the import of a reduction as a consequence of which a set of distinctive traits of some subject matter is assimilated to what is patently a set of quite dissimilar traits. In such cases, the distinctive traits that are the subject matter of the secondary science fall into the province of a theory that may have been initially designed for handling qualitatively different materials and that does not even include some of the characteristic descriptive terms of the secondary science in its own set of basic theoretical distinctions. The primary science thus seems to wipe out familiar distinctions as spurious, and appears to maintain that *what are prima facie indisputably different traits of things are really identical* [emphasis mine]. The acute sense of mystification that is thereby engendered is especially frequent when the secondary science deals with macroscopic phenomena, while the primary science postulates a microscopic constitution for those macroscopic processes". See also Nagel 1970: 368 "Although much scientific inquiry is directed toward discovering the determining conditions under which various traits of things occur, some of its important achievements consist in showing that things and processes initially assumed to be distinct are in fact the same". Although these suggestions, however, Nagel explicitly and stably states that the term reduction applies to theories and statements, remaining coherent with his semantic approach. In the 1970 paper, for instance, he confirms that "Scientists and philosophers often talk of deducing or inferring one phenomenon

of postulates, theorems, statements and laws.

2.4.1 *Nagelian and post-nagelian models of reduction*

Ernest Nagel developed his model of theory reduction for years. The first formulation of the theory traces back to 1949,[201] and was then developed in the 1961 book *The Structure of Science: Problems in the Logic of Scientific Explanation*. Here, he emphasises that one of the peculiar features of science is “the phenomenon of a relatively autonomous theory becoming absorbed by, or reduced to, some other more inclusive theory”.[202] Nagel calls “secondary science” the reduced or absorbed (set of) theories, and “primary science” the reducing, absorbing one. He described, furthermore, two different kinds of theory reduction: the first is a straightforward, “homogeneous” reduction holding between theories explaining with a similar conceptual apparatus similar phenomena taking place into different domains. For instance, from Aristotle’s *Physics* until Eighteen century, terrestrial and celestial motions were supposed to follow different mechanics, but with the formulation of Isaac Newton’s three laws of motion these two theoretical frameworks were reduced to a single one, and terrestrial and celestial phenomena revealed themselves as different manifestations of the same natural principles. In this case, the reduction at issue is *homogeneous* as neither logical dilemmas, nor bridge laws, nor new special vocabularies are required for the reduction (i.e. the explanation) of the secondary theory to (by) the first.[203] In Nagel’s words: “[…] the laws of the secondary science employ no descriptive terms that are not also used with approximately the same meanings in the primary science. Reductions of this type can therefore be regarded as establishing deductive relations between two sets of statements that employ a homogeneous vocabulary”.[204]

By contrast, the second kind of reduction, the *inhomogeneous* one, is described by Nagel as a very complex process, where the secondary and primary sciences use different

---

from another [...] However, these locutions are elliptical, and sometimes lead to misconceptions and confusions. For strictly speaking, it is not phenomena which are deduced from other phenomena, but rather statements about phenomena from other statements” (Nagel, 1970: 360).

[201] Nagel 1949.

[202] Nagel 1961: 336-7.

[203] Even this straightforward kind of reduction, however, is rare. See Sklar 1967: 110 “Matters are not, however, quite this simple even in these very rudimentary cases of reduction. If one looks for examples of reduction from the history of science, strictly derivational reductions are few and far between. One can construe various relationships of a strictly derivational sort as reductions, but an examination of cases of reduction pre-analytically so-called, shows that even in the case of homogeneous theories reduction is very rarely derivation”.

[204] *Ivi*: 339.

vocabularies and describe dissimilar phenomena in dissimilar ways. A paradigm instance of this second model is the debated reduction of thermodynamics to statistical mechanics, which can be briefly portrayed as follows.

Thermodynamics[205] (in the classic sense) is the study of continuous portions of matter - defined as *control volumes* - separated from the surrounding environment by a more or less insulating and mobile boundary - defined as a *control surface*. These portions of matter, moreover, being immersed in an environment, can be open systems - that is, they can be subject to flows of matter and energy - and thermodynamics examines precisely these energy exchanges that occur in the so-called *thermodynamic universe* (which is precisely the whole made up of "system + environment"). We have said that control volumes are *continuous* portions of matter and this is a technical term that indicates how thermodynamics describes the systems it studies from a macroscopic point of view. In physics, that is, matter can be described from a microscopic (atomic or molecular) perspective or from a macroscopic, that is, continuous perspective. In the first case, the object of study is all the information that the analysis of every single atom or molecule in the system can provide. This amount of information, obviously, is enormous if we consider that a thermodynamic system is microscopically constituted by a number of elements in the order of Avogadro's number (which is equal to 6.022×10^{23}).[206] In the case of thermodynamic systems, however, this disproportionate amount of information is rather useless, as the macroscopic behaviour of the system is insensitive to the position and speed of all its individual components. This is why thermodynamics adopts a continuous and macroscopic perspective: matter is considered as a continuum and the information useful to describe it is simply the average of all microscopic information.

A common example of a thermodynamic system is the one we all studied in school, namely a gas inside a container. In this case the gas is the control volume and the container is the control surface. This thermodynamic system is described by the values of some macroscopic properties which are called thermodynamic variables or state functions. These include, for example, pressure, volume and temperature and if these variables are constant, the system is in equilibrium, while if they are not, it is not in equilibrium and energy flows occur. The behaviour of the thermodynamic system follows the laws of thermodynamics and is described, as said, by the relative variables. Thermodynamic laws and variables are

[205] See Sonntag, Borgnakke, Van Wylen, & Van Wyk 1998.
[206] See Mazzoldi, Nigro, Voci 2003: 364 and ff.

macroscopic and this means that the thermodynamic state of a system is defined by such macroscopic quantities, therefore it does not have a conceptual correspondence with its mechanical state which is instead defined by the knowledge of the position and speed of all the elements from which the system is composed: "if the thermodynamic state is known, the mechanical state is not generally known; on the contrary, a given thermodynamic state can correspond to many different mechanical states".[207] This happens because the macroscopic behaviour of a thermodynamic system is partially independent from its exact microscopic composition and depends on other factors such as the extension of the system and the properties of the surrounding environment.

Statistical mechanics, on the other hand, is a branch of mechanics that introduces statistical methods and probability theory. It applies when the initial conditions of a system or its laws are not perfectly known, or when the system is open. In other words, statistical mechanics is used in all those cases where classical mechanics is not able to make precise predictions about the behaviour of a physical system. Statistical mechanics uses probability theory to provide predictions about systems made up of a large number of components, and these predictions are statistical, so they do not concern the exact state of each individual component of the system (such as the position and speed of each of its atoms), but its general behaviour. As we will see in §3.2.1 speaking about information, probability cannot foresee the precise result of tossing single coins, for example, but it can foresee how often the result will be head or tails. Statistical mechanics can therefore describe the macroscopic behaviour of systems composed by a great number of atoms or molecules such as thermodynamic systems.

Thermodynamics and statistical mechanics therefore share some notions, such as volume or pressure, while other thermodynamic concepts, such as temperature, heat or entropy, are not found in the theoretical framework of statistical mechanics because they are macroscopic and observational notions. If we want to reduce the secondary theory to the primary one, however, an essential task is to determine and understand the relations between the concepts of the former and the latter: a task often difficult to accomplish for various reasons,[208] as we will see shortly.

Having drawn the distinction between homogeneous and inhomogeneous reductions, and provided some examples of them, Nagel expresses some formal conditions, which can be

[207] Ivi: 365.
[208] Si veda Sklar 1999.

summarized in the idea that in the case of a secondary theory supposed to be reduced to a primary theory, the fundamental theoretical statements of the secondary discipline (Nagelian "T statements") must be *logically deducible* from the theoretical postulates of the primary theory.[209] This model required, moreover, that all the involved statements are explicitly formulated, that their meaning is fixed, and that the primary and secondary theories share theoretical expressions and terms having approximately the same meaning. As far as inhomogeneous reductions are concerned, however, this last requirement is notoriously critical, and justifies the necessity of additional assumptions postulating "suitable relations" or "linkages" between critical terms of the reduced theory and some correspondent statements or entities of the reducing theory (for instance: the linkage between the notion of temperature in thermodynamics and the notion of mean molecular kinetic energy in statistical mechanics). We will see, however, that in the case of this reduction, the matter is much more complex. Starting from Nagel's 1970 paper *Issues in the Logic of Reductive Explanations*, these assumptions or linkages have been called "bridge laws", and in this last paper Nagel admits that they are often partial similarities or approximations, rather than sweeping logical derivations.[210]. Nonetheless, Nagel's model of reduction has been widely criticised as too narrow and unrealistic, and therefore modified by many philosophers, such as Paul Feyerabend,[211] Lawrence Sklar,[212] Kenneth Schaffner,[213] and Patricia Churchland, who in the following quotation sharply identifies different varieties of reductions:

> At one end of the spectrum then, we have pairs of theories where the old is smoothly reduced by the new, and the ontology of the old theory survives, although redescribed, perhaps, in a new and more penetrating vocabulary […] In the middle of the spectrum, we find pairs of theories where the old ontology is only poorly mirrored within the vision of the new, and it "survives" only in a significantly modified form. Finally, at the other end of the spectrum we find pairs where the older theory, and its ontology with it, is eliminated entirely in favor of the more useful ontology and the more successful laws of the new.[214]

[209] See Nagel 1961: "[…] a reduction is effected when the experimental laws of the secondary science (and if it has an adequate theory, its theory as well) are shown to be the logical consequences of the theoretical assumptions (inclusive of the coordinating definitions) of the primary science".
[210] See Nagel 1970: 362: "[…] in actual scientific practice, the derivation of laws from theories usually involves simplifications and approximations of various kinds, so that even the laws which are allegedly entailed by a theory are in general only approximations to what is strictly entailed by it".
[211] Feyerabend 1981.
[212] Sklar 1967 and 1993.
[213] Schaffner 1993.
[214] Churchland & Churchland 1992: 23

A single model of inter-theoretical reduction therefore seems difficult to define, and the same goes for the nature of the bridge laws and the ontological status of the entities concerned by the theories involved in the reduction. As also highlighted by Sklar and Schaffner, for example, the reduction process often involves a *correction* of the reduced theory rather than an actual reduction to the most fundamental theory. Schaffner takes this problem into account when formulating his version of inter-theoretical reduction in 1977,[215] i.e. the model called "General Reduction Replacement" (GRR).

This model is a revised version of the Nagelian reduction,[216] and what Schaffner wants to highlight is that the reduction of an older theory, called T1, to a newer theory called T2, corresponds, on closer inspection, to the production of a further revised theory, i.e. T1*, which although similar to T1 is not equal to it, but is the theory that is actually explained by T2. In other words: T2 explains T1* and not T1, but T1* is very similar to T1, and the latter circumstance allows us to state that, broadly speaking, (also) T1 has been reduced to T2. This process, however, is still an approximation that does not correspond to a totally exhaustive reduction of the original theory.

A clear example of this is the alleged reduction of thermodynamics (which would represent T1) to statistical mechanics (T2), a reduction that actually produced statistical thermodynamics (T1*),[217] which is the study of how the macroscopic and thermodynamic behaviour of a system can be related to the statistical state of its microscopic components. It should also be noted that mechanics and thermodynamics present incompatible principles highlighted by two paradoxes, that of reversibility and that of recurrence.[218]

Now, reconciling two disciplines that present incompatible principles is not a linear operation and therefore requires a reformulation or at least a reinterpretation of those theoretical assumptions that underlie them. In fact, mechanics and thermodynamics can only become compatible at the cost of a "less rigid interpretation of the notion of equilibrium". We will come back to this point later on, at the end of the chapter, because these revisions of the theoretical principles of a discipline play an important role in decreasing the incompatibility of reduction and emergence. In sum, as we have said, there is a first general model of reduction for which the secondary theory T_2 is reduced to the

215 Schaffner, 1977.
216 See Schaffner, 2012: 540 and ff.
217 For more on this, see Sklar 1993.
218 See Falcioni, Vulpiani 2015: 1 ff..

primary theory T_1 if it is logically deducible from and explainable by it, with or without correlational bridge laws or statements. There is, moreover, a second model for which the secondary theory T_2 is in some sense *replaced* by a strongly analogous theory T_2^* which is reduced to or explained by the primary theory T_1, and this analogy allows to state that the original T_2 has been reduced to T_1. Further to this, we have at least another model, in which no bridge laws or replacements are involved, and this case is exemplified by the reduction of classical mechanics (CM) to special relativity theory (SRT).

2.4.2 *Reduction in the limit*

As already mentioned in footnote 122, in this case is not at all clear which of the two theories is the most fundamental and in which direction the reduction should be conducted. Let's adopt the position for which CM can be reduced to SRT. How is it possible? As stated by Sklar:

> If we first restrict our attention to sentences framable in purely kinematic concepts, and then further restrict our attention to the subset of sentences in this class dealing with sufficiently low velocities, we find for every sentence in this subset derivable from the relativistic theory there is a sentence approximative to it derivable from the Newtonian theory.[219]

For instance, following Nickles,[220] if we focus on momentum, we will discover that in a particular limit the Einsteinian relativistic formula reduces to the classical formulation. Let me offer a brief digression which will clarify the question.

The classical formula of momentum states that the momentum (p) of a body is directly proportional to the mass (m) and the velocity (v) of the body:[221]

$$p = mv \tag{1}$$

Nonetheless, in the case of relativistic speeds, namely with velocities comparable to that of light (c), the formula (1) is modified through the addition of the factor γ, so that the

[219] Sklar 1967: 116.
[220] Nickles 1973.
[221] The standard unit of momentum is therefore kilogram-meter per second (kg · m/s)

relativistic momentum, which simply means the momentum of a body moving at relativistic speeds, can be formulated as

$$p = \gamma m_0 v \quad (2)$$

where m_0 is the rest mass and the relativistic factor γ is equal to

$$\gamma = \frac{1}{\sqrt{1-v^2/c^2}} \quad (3)$$

Now, from (2) and (3), we obtain the extended formula of the relativistic momentum, which is the following:

$$p = \frac{m_0 v}{\sqrt{1-v^2/c^2}} \quad (4)$$

To sum up, we have a classical mechanics formulation of the momentum (1), and a relativistic formulation of it, (4). The keystone of the reduction of (4) to (1) is the value of the relevant parameter v, i.e. the velocity of the body. If the value of v tends to zero ($v \to 0$), the ratio between v and the speed of light c will be negligible[222] making γ close to 1, hence neutral to the product (2). In this case, therefore, the relativistic equation (4) will be reducible to the classical formula (1) without the relativistic factor γ. This circumstance is pretty obvious, because γ is necessary in the limit $v \to c$, that is with values of v tending to the speed of light, while in the opposite case, with values of v tending to zero, it can be neglected.

This notion of "reduction in the limit" is very common in physics[223] and involves no bridge-laws or substitutions of any kind because the theory that is reduced seems to be a particular case of the one that should reduce it, rather than a different theory correlated in some way with the more fundamental one. What we have taken as an example, however, is not a special or anomalous case: as clarified by Chibbaro, Vulpiani and Rondoni, philosophers tend to conceive reduction as a process that leads less fundamental macroscopic

[222] Let's make an example with $v = 10^3\ m/s$. This is a relevant speed, macroscopically speaking, as it is equal to 3600 km/h, which is abundantly supersonic (the speed of sound is 1,235 km/h). With $v = 10^3\ m/s$, and given the speed of light, which is equal to $3\times10^8\ m/s$, the ratio v^2/c^2 contained in γ will be $10^6/9\times10^{16}$, which, simplified, is equal to $1/9\times10^{10}$, namely 0,00000000001111. The denominator of the factor γ will be therefore $\sqrt{1-0{,}00000000001111}$, which can be approximated to 1, being it 0.999999999994445. Being equal to 1, the relativistic factor is neutral and the formula for the relativistic momentum $p = \gamma m_0 v$ becomes equivalent to the classical formula $p = mv$. This equivalence means that in the limit $v \to 0$, γ approximates to 1, becoming negligible.

[223] For a review, see Chibbaro, Rondoni e Vulpiani 2014: 32.

theories to be deduced from and explained by more fundamental theories that usually affect the microscopic level of reality. Instead, in physics reduction is conceived in opposite terms, i.e. as a process that leads to the absorption of less general and therefore more fundamental theories by those theories that manage to explain a larger number of phenomena, both microscopic and macroscopic.

The example of Newton mechanics and special relativity is clear: the elaboration of the theory of relativity does not lead to the abandonment or revision of Newton mechanics, but on the contrary it can itself be reduced to the former in the limit, that is, in the specific case in which the speed of the body in question is close to infinity. This model of reduction used in physics therefore does not suggest that the progress of science will lead to the abandonment of all the "old" and "macroscopic" theories, declaring them inadequate, nor that their ontologies will be declared as overcrowded with fanciful entities that will have to be abandoned in favour of others. This model shows, on the contrary, that in many cases inter-theoretical reductions correspond to *integrations* and *reconciliations* between different theories and ontologies related to different levels of organization of reality, rather than to the elimination of some of them in favour of others.

2.5 *Conclusions*

Over the twentieth century different types of inter-theoretical reduction have been developed, but their nature, which exploits approximate and ideal models, suggests that, as a rule, there is no real identity or replaceability between the old reduced theory and the new theory that reduces it. In short, borrowing Schaffner's expression, we notice a significant divergence between ideal reductions that are simple, linear and definitive (sweeping), and real reductions that are instead cumbersome, slow and gradual (creeping) and can only be carried out following *ad hoc* theoretical reformulations and adjustments. For this reason, as in the case of ontological reduction, when intertheoretical reduction is completed, we realize that something has remained excluded and that the reductive process could not explain every aspect of the theories or sciences that had to be reduced. At this point, the question arises, again, whether the emergence cannot affect those residual parts that escape intertheoretical reduction. To explore this hypothesis it is useful to return to the question of the reduction between thermodynamics and statistical mechanics.

We have said that these two disciplines have incompatible principles that create paradoxes regarding the concepts of recurrence and irreversibility. Let us now quickly see what these principles are and what these paradoxes entail. As we have seen, thermodynamics studies portions of matter isolated from a surrounding environment and these thermodynamic systems may or may not be subject to exchanges of matter or energy. When a system is not subject to any exchange of energy, it reaches a state of equilibrium that it cannot spontaneously abandon. Imagine two bodies in contact. If the two bodies are one colder and the other warmer, energy exchanges will take place until the two objects reach the same temperature. At that point, the system will have reached a state of equilibrium different from the initial state of non equilibrium and the energy exchanges will cease. Without external intervention, the system will no longer abandon the state of equilibrium and this is the circumstance that justifies the introduction, in thermodynamics, of the notion of *irreversibility*: a process is irreversible when it does not allow an inverse temporal evolution. Our two objects, for example, will not spontaneously return to a state of non-equilibrium in which one is hot and the other is cold unless someone or something intervenes to heat one or cool the other. The process that leads from an initial state of non-equilibrium to one of equilibrium is therefore irreversible, and the notion of irreversibility, linked, as we know, to that of entropy, is central in thermodynamics. Yet, irreversibility represents a problem for the conciliation between thermodynamics and mechanics, since mechanical systems are always reversible and temporally symmetrical (the equations of motion are symmetrical with inverted time).

This characteristic of mechanical systems is described by the Recurrence Theorem formulated by Poincaré, which states that certain systems will, after a sufficiently long but finite time, return to a state arbitrarily close to (for continuous state systems), or exactly the same as (for discrete state systems), their initial state.[224]. Therefore, according to Poincaré's theorem, any system, including the one formed by the two objects that have reached a state of thermal equilibrium, will sooner or later return to a state of non-equilibrium close to the initial one. How, then, can we reconcile Boltzmann's theorem of irreversibility with Poincaré's? This question is equivalent to asking how to reconcile the microscopic and temporally symmetrical world of mechanics with the macroscopic and temporally

[224] Falcioni, Vulpiani 2015: 57. Poincarè's theorem reads: "Given a Hamiltonian system in a limited phase space Ω , and a set $A \subset \Omega$, all trajectories starting from $x \in A$ will return to A after a certain time and this will happen infinitely, except for initial conditions in a null set". (Ivi: 121).

asymmetrical world of thermodynamics, a problem that represents the core of the reduction in question.

An intuitive way to untangle this issue is to think about what the recurrence of Poincaré's theorem corresponds to. The theorem says that in every system, every dynamic state will arbitrarily repeat itself. The question to ask, at this point, is *when*. How long will it take a state to return, i.e. to recur? Poincaré's theorem does not address this problem, which is answered instead by Kac's lemma.[225] The recurrence time depends on the amount of elements that make up the system. A system with very few elements will have few states that will alternate in a relatively short time, while a system with many elements will have more states and it will take more time for them to recur. In other words, the recurrence time increases with the number of system elements and, consequently, with the increase of the variables needed to describe it. By quantifying these elements more precisely, however, we will realise that a system, for recurrence times to be moderate, must be composed of a *tiny* handful of elements. If we hypothesize a system composed of a hundred or so elements (10^2), for example, the recurrence time will already be equal to the age of the universe.

Each thermodynamic system, however, is made up of a disproportionately higher number of elements than Avogadro's, i.e. 6.022×10^{23}. The recurrence time, for systems of this type, can therefore be considered infinite, and the probability that the system will return to the initial state of non-balance is, in turn, infinitesimal, that is, negligible. In other words: "Recurrence must be there because a rigorously demonstrated theorem requires it, but we discover that for macroscopic systems, which are the object of thermodynamics, the typical time to observe it, on which the theorem remains silent, is enormously long even adopting astronomical time scales".[226] What seemed to be a paradox, i.e. the incompatibility between the reversibility typical of systems described by mechanics and the irreversibility of thermodynamic systems can therefore be reduced by taking into account the different variables involved at microscopic and macroscopic levels. In considering the question it is therefore necessary to keep in mind, in addition to the time scales, also the different levels of reality involved, which become evident in the conceptual experiment elaborated by Falcioni and Vulpiani:

> […] pour perfume in the corner of a room; the molecules of the perfume, initially concentrated in a small region, will quickly occupy the whole space. Now imagine that you

[225] Ivi: 122.
[226] Falcioni, Vulpiani 2015: 14.

> can film the molecules. By playing the film backwards, you will see an unnatural phenomenon: all the molecules scattered in the room will gather in a corner. If you look at just one molecule you will not see anything abnormal in the film played backwards. Similarly, there is nothing strange in the film played backwards if one limits observation to the molecules contained in a small area of the room over a not-too-long period of time. The impression of unnatural behaviour can only be gained by looking at a large number of molecules or a fairly large area of the room.[227]

In other words, irreversibility is not something that one can talk about with reference to individual molecules or the individual elements that make up a system. The motion of the molecules taken individually is always reversible, while the collective behaviour of the system formed by a large number of molecules is not. Irreversibility is therefore a phenomenon that depends on the level of organisation of matter: it occurs at collective macroscopic levels, as in thermodynamic systems, while at the microscopic level it simply does not exist. In this sense, one can say that irreversibility is an emergent property of certain systems that are in a particular condition, defined by two circumstances: (i) being composed of a number of elements that tends to infinity, (ii) whose dimension tends to zero.[228] And these are the systems that, on closer inspection, make up the entire macroscopic world. These systems are as real as their properties are and they do not depend on the type of observation carried out or on our knowledge of them. A system is not irreversible *because* it is observed from the outside using probabilistic instruments. Irreversibility is a physical fact that occurs under certain conditions linked to a particular level of organisation of matter.

To return to the question of the relationship between epistemological irreducibility and emergence, it will now be clear that the matter is more complex than it seemed. On the one hand, it is not accurate to say that emergent phenomena cannot be exhaustively reduced at an inter-theoretical level since this definition applies to all phenomena in general and therefore lacks precision: no phenomenon can be exhaustively reduced to another, no particular theory explains exactly all phenomena explained by a more general (less fundamental) theory, and cases of satisfactory inter-theoretical reduction are rare and linked to the implementation of particular measures.[229] If, therefore, no phenomenon can be effectively reduced to another on an epistemological level, saying that emergent phenomena are epistemologically irreducible

[227] Ivi: 67.
[228] On this, see Falcioni & Vulpiani 2015.
[229] Schaffner 2006: 378.

means not saying much (although it is not false). On the other hand, that the reduction is not ideal and complete does not imply that it is useless or ineffective. The reductive method has enabled the progress of science and it is essential to recognise all its qualities. If we look at the reductions that are made in the natural sciences, however, we see that they are always accompanied by the introduction of new semantics to describe high-level phenomena which, although reduced, maintain their autonomy and specificity. Emergence, therefore, does not seem to be ruled out by the presence of effective (albeit partial) inter-theoretical reductions: each reduction leaves something irreduced and must be supplemented by the formulation of new concepts that correspond to high-level natural phenomena. These, in fact, cannot be exhausted or described by the conceptual apparatus of the low-level theory to which the former is reduced. It is therefore clear that both at a conceptual and an ontological level, effective reductions do not exclude emergence, but imply and validate it, as happens for irreversibility in the reduction of thermodynamics to statistical mechanics: a reduction that at this point will have to be seen rather as a conciliation or integration.

What, then, of the thesis that a phenomenon must be irreducible to be emergent? Important considerations remain to be made on this point. First of all, we have seen that both in the case of ontological reduction and in that of epistemological reduction, reductionism does not keep the promises that were (perhaps too optimistically) attributed to it.[230] Reduction is effective and useful, but it is never absolute and universal. It is a precious methodological tool which, however, is not to be attributed a totalizing metaphysical scope. Reduction never proves that a certain range of phenomena do not *actually* exist. It proves that some of their properties - but not all of them - are realized by the properties of their constituents (ontological reductionism) and that part of the knowledge we have of them - but not all of them - can be deduced from other knowledge, sometimes related to more fundamental phenomena (epistemological reductionism). In both cases, however, reductionism intrinsically implies the presence of something residual that, in my opinion, can be reasonably understood as an emergent phenomenon, be it a part of reality made up of those properties that are neither functional nor directly causal, or a part of the semantics through which we describe the world. Therefore, to think that there are phenomena for which reduction *fails* does not accurately describe the dynamics we have reconstructed in these pages. Ontological reduction works when it is applied to the right range of phenomena, i.e. those that can be defined in functional or causal terms. If the reduction fails it is because the reductive method has been applied to the wrong phenomena.

[230] See Dirac 1929; Reichenbach 1959; Putnam and Oppenheim 1958; Feynman et al. 1964; Nagel 1979.

Therefore, it is not the reduction itself that fails, but the desire to transform it into a universal tool.

Secondly, claiming that emergent phenomena are irreducible seems to suggest that where there is emergence there is no reduction and vice versa, but this suggestion takes us a long way from where we would like to go. Emergence and reduction are strictly complementary. One implies the other because there is no high-level phenomenon that is not partly reducible and partly emergent. Some phenomena are almost completely reducible and others almost completely emergent, but reduction and emergence cannot be clearly separated and this depends on their very nature. On the one hand, reduction is a simplification and any simplification implies that a series of relevant features are maintained at the expense of others; on the other hand, emergence is never the appearance of something from nothing (otherwise it would be dualism), but it is always an emergence from something, and that something is a more fundamental phenomenon that certainly has features that can be treated in reductionist terms. Having made this clarification on the relationships between reduction and emergence (which are of complementarity and not of clear opposition), a second and final consideration is now necessary.

We have repeatedly anticipated a pluralist stance towards emergence and we have also mentioned an approach that makes use of the notion of an open property cluster. Within this cluster we mentioned the irreducibility of emergent phenomena, an irreducibility that, at this point, should be redefined in terms of complementarity. Emergent phenomena, which are high-level phenomena exhibiting emergent properties, also instantiate, in other words, non-emergent properties, even if the instantiation of the former has been seen as the peculiarity able to make the phenomena in question generally emerge. Our own open cluster must therefore include this complementarity, which should be understood as a mutual completion, where recognising the presence of reducible and emergent traits is the best way to provide a complete description of any high-level phenomenon. In short, our cluster includes *the property of simultaneously instantiating reducible and non-reducible properties* - (at an ontological and epistemological level, and with different meanings depending on the type of reduction analysed), respecting the complex and multidimensional character of any high level phenomenon.

CHAPTER III

Novelty

3.1 *Introduction*

As we have seen in previous chapters, emergence theorists agree that emergent phenomena exhibit new characteristics, new causal powers and new behavioural patterns, as well as requiring the elaboration of new concepts and theoretical frameworks. In other words, these phenomena exhibit *ontological and epistemological novelty*. However, in most cases, epistemological novelty derives from ontological novelty, as already pointed out by the British Emergentists,[1] so in this chapter we will explore the ways in which ontological novelty has been interpreted by different emergence theorists.

At the beginning of the debate, for example, John Stuart Mill talked about "heteropathic effects" produced by phenomena governed by new laws, and related novelty to *non-linearity*[2] and *unpredictability*. These two features are still considered two hallmarks of emergent entities, which in this perspective correspond to composite entities whose

---

[1] See McLaughlin 1992: 55 (footnotes) "Emergentists often speak of emergent properties and laws as unpredictable from what they emerge from. But, contra what some commentators have thought, the Emergentists do not maintain that something is an emergent because it is unpredictable. Rather, they maintain that something can be unpredictable because it is an emergent. Emergence implies a kind of unpredictability. But it is a mistake to conflate emergence with this consequence of emergence. The British Emergentists do not".

[2] I will develop an analysis of the concept of non-linearity in paragraph 3.2., but a provisional definition may be useful from the outset, given the importance of the concept highlighted here. Non-linearity is the property of systems composed of constituent parts that mutually influence each other generating complex relational dynamics. These systems do not respect the so-called *overlapping principles*, i.e. the principles of homogeneity and additivity. The first states that in a system the output is directly proportional to the input, while the second states that the output of the whole must be exactly the sum of the outputs of the parts taken individually. These principles are respected by so-called linear systems, but not by non-linear ones, which, given the correlation between the variables that characterize their internal dynamics, show unpredictable properties and behaviours.

constituent parts interact with each other in a non-additive way, generating the appearance of properties at the systemic level that the parts at lower levels of organisation do not possess. For these reasons, the properties that characterize an emergent phenomenon cannot be deduced prior to their first instantiation and observation, and this circumstance shows how ontological novelty leads to another epistemic one. We deal with non-linearity in § 3.2.

A few years later and with reference to the same phenomena, novelty has been described as *bruteness*: something is authentically new when it is a brute fact, i.e. a primitive, basic phenomenon that cannot be further analysed or explained. This characterisation can be traced back to British Emergentists such as Alexander and Broad, but is still used by some contemporary scholars such as Jessica Wilson, Mark Bedau, Elisabeth Barnes and John Symons. §3.3 is dedicated to these themes.

Another sense of novelty was developed within the philosophy of evolution in the 19th century, when some excellent scholars such as Thomas Huxley and the British Emergentist scientist Conway Lloyd Morgan noted the importance of integrating Darwin's gradualist theory of evolution with what they called *qualitative novelty*. Darwin's attention focused on the accumulation of quantitative changes and the continuity of the evolutionary process, while these authors recognised the importance of admitting a sort of qualitative novelty that consisted in the heterogeneity observable in nature: a heterogeneity that the mere accumulation of slightly different and gradual quantitative changes could hardly explain. This is the subject of § 3.4, in which we analyse another case of emergent phenomenon involving qualitative novelty as heterogeneity: this phenomenon is spacetime.

What has been considered the most relevant type of novelty, however, is the *causal novelty*, to which this dissertation dedicates an entire chapter (chapter four). Causal novelty has long been considered the glaring mark of ontological emergence because the authors who dealt with emergence attributed much importance to the presence of new causal powers, often capable of influencing the underlying phenomena from which high-level phenomena emerged. Let's now examine all these different kinds of novelty.

3.2 *Heteropathic effects and nonlinearity*

It might be pointed out that together with nonlinearity, fundamentality, qualitative and causal novelty, also unpredictability (i.e. epistemological irreducibility) can be considered as a kind of novelty. However, as already mentioned, unpredictability is a consequence of novelty and not its cause or something that can be considered correspondent to it. McLaughlin noticed that the British Emergentists have highlighted this point since the very beginning of the debate about emergence, and we agree with them:

> Emergentists often speak of emergent properties and laws as unpredictable from what they emerge from. But, contra what some commentators have thought, the Emergentists do not maintain that something is an emergent because it is unpredictable. Rather, they maintain that something can be unpredictable because it is an emergent. Emergence implies a kind of unpredictability. But it is a mistake to conflate emergence with this consequence of emergence. The British Emergentists do not.[3]

Despite never having used the word "emergence", John Stuart Mill is considered the father of British Emergentism because of his attention to certain natural phenomena he considered inexplicable by the scientific knowledge of the time. In the third book of *A System of Logic*, "Of the Composition of Causes", Mill distinguished between *homopathic* and *heteropathic* effects reporting the essential difference, in nature, between

> The case in which the joint effect of causes is the sum of their separate effects, and the case in which it is heterogeneous to them; between laws which work together without alteration, and laws which, when called upon to work together, cease and give place to others.[4]

Mill begins his analysis focusing on what in mechanics is called the *Principle of Composition of Forces*, which states that the joint effect of two forces acting together is equal to the sum of their separate effects as taken in isolation. This is Mill's example:

> If a body is propelled in two directions by two forces, one tending to drive it to the north, and the other to the east, it is caused to move in a given time exactly as far in both directions as the two forces would separately have carried it; and is left precisely where it would have arrived if it had been acted upon first by one of the two forces, and afterwards by the other.[5]

[3] McLaughlin 1992: 55 (footnote 31).
[4] Mill 1843: 212.
[5] *Ivi*: 210.

In imitation to this law, Mill talks about *Composition of Causes*, a principle holding when "the joint effect of several causes is identical with the sum of their separate effects".[6] Nonetheless, in some natural systems, such as the chemical or biological ones, this last principle fails and the conjoint effect of this kind of causes cannot be reduced to a vector sum:

> Not a trace of the properties of hydrogen or of oxygen is observable in those of their compound, water. The taste of sugar of lead is not the sum of the tastes of its component elements, acetic acid and lead or its oxide; nor is the color of green vitriol a mixture of the colors of sulphuric acid and copper.[7]

In certain cases, moreover, "[…] the concurrence of causes is such as to determine a change in the properties of the body generally, and render it subject to new laws, more or less dissimilar to those to which it conformed in its previous state of existence".[8] These last cases exhibit what Mill defined as *heteropathic* effects, which are regulated by *heteropatic* (novel) laws. Heterogeneity descends from the fact that the laws governing a phenomenon when it is isolated are different from the laws governing the same phenomenon when it is in combination with other phenomena. *New* laws appear and govern the organised entity *interfering* with the laws already existent and governing the components, and these novel laws "may supersede one portion of the previous laws but coexist with another portion, and may even compound the effect of those previous laws with their own".[9] What is emergent in Mill's theoretical framework, therefore, are laws and relationships between causes and effects, and the emergence of these new laws and relations makes the dynamics of the systems in which they appear *unpredictable*. Eventually, unpredictability (together with the appearance of new laws and features) is the feature characterising heteropatic phenomena: the phenomena that starting from Lewes[10] will be called *emergent*.

In more recent times, the nature of what Mill called heteropatic effect – i.e. nonlinearity – has been widely studied in science and philosophy. The novelty ascribed by Mill to emergent phenomena, therefore, can be correlated to this feature, and this circumstance is

[6] *Ivi*: 211.
[7] Ibidem.
[8] *Ivi*: 435
[9] Mill 1843: 213.
[10] Lewes 1877.

significant, for today nonlinearity is recognised as a central feature of complex systems, which are well-known cases in which emergence is (supposed to be) present.

In the philosophical literature, Mark Bedau and Jessica Wilson are two examples of scholars who focus on nonlinearity[11] wondering whether it can still be considered one of the marks of emergence. As we will see, Bedau concludes that nonlinearity – and in particular underivability and incompressibility – is correlated with a metaphysically innocent form of emergence he defines *weak emergence*. On the other hand, Wilson, who focuses on metaphysical accounts of emergence, concludes that nonlinearity is not a sufficient criterion for ontological emergence because it is too inclusive and characterises phenomena that do not involve any novel or fundamental emergent feature.[12] Let's now briefly outline what is nonlinearity and how and why it can be considered a mark for a certain kind of emergence but not for others. Obviously, within the present pluralist framework, non-linearity, despite not being a necessary and sufficient condition of emergence, is part of the open cluster of properties that describes emergent phenomena. The cluster, as we have seen, does not require absolute and immutable conditions, but is limited to listing the characteristics that emergent phenomena seem to usually manifest: non-linearity, as we will see shortly, is certainly one of them.

3.2.1 *Linearity and nonlinearity*

The distinction between linear and nonlinear systems is of central importance in science. The terms *linear* and *nonlinear* origin from the shapes of the graphs of (linear and nonlinear) equations. As is shown in Figure 3.1, the diagram of $y = x$, an easy to be solved polynomial equation of degree one, is a straight line, while the graph of $y = x^2$, a more complex equation of degree more than one, displays a quadratic curve. Linear equations, although mathematically simpler, can only represent a certain number of phenomena, as the world we live in is only partially linear. Nonlinear equations are essential for modelling nonlinear events, but they are harder – often impossible – to be solved, so, in many cases, mathematicians try to *linearize* nonlinear models, which means they try to approximate nonlinear models to linear ones. However, these approximations only work in limited conditions, and can only describe local behaviours as well as elementary dynamics.

[11] Bedau 1995, Wilson 2013 and 2019.
[12] See Wilson 2013.

As far as systems are concerned, a linear system (that could be described by linear equations) respects the so-called *superposition principles*, which are the homogeneity and the additivity principles. The homogeneity principle states that *for every linear system, the net output is directly proportional to the input*, while the additivity principle states that *the output of the whole is exactly the sum of the outputs of the parts as taken in isolation* (and here the similarity with the principle of composition described by Mill is neatly evident). The reason why certain systems are linear and other are nonlinear depends upon the way in which their variables are (or are not) related. Linear systems exhibit *variables uncorrelation*, which means that the dependence and covariance of two random variables are equal to zero.[13] By contrast, nonlinear systems show *variables correlation* in that variables influence each other creating more complex dynamics and "an intricate graph of causal links".[14] Ontologically speaking, therefore, nonlinear systems are systems in which in addition to the intrinsic properties of the components, the relationships between them play a relevant role for the system, while linear systems are cases in which the only relevant properties are the ones intrinsic to the components. From an epistemological point of view, instead, nonlinear systems are characterised by unpredictability, as stated by Mill as well. Because of their

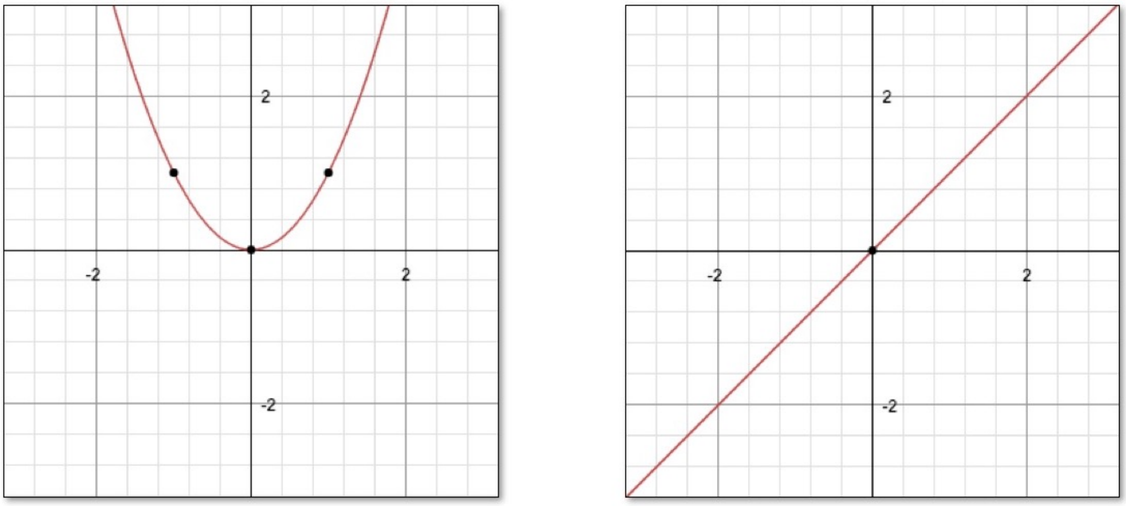

Figure 3.1 Graphs of the equations y = x, on the right, and $y = x^2$, on the left. The first is a straight line, from which the term "linear equation", while the second is a quadratic curve.

[13] See MacKay 2008: 274 "superposition means that the joint probability distribution is a product, that is the particles behave independently". See also Bishop 2010: 113-114: "Linear systems can be straightforwardly decomposed into and composed by subsystems (a consequence of the principle of linear superposition) […] The linear behavior of such systems in these cases is sometimes called resultant (in contrast with emergent). In nonlinear systems, by contrast, this straightforward idea of composition fails (a consequence of the failure of the principle of linear superposition). When the behaviors of the constituents of a system are highly coherent and correlated, the system cannot be treated even approximately as a collection of uncoupled individual parts".
[14] MacKay 2008: 273.

sensitivity to initial conditions and to variables correlation, any measurement inaccuracy, however minimal, leads to enormous errors in prediction, so it is impossible even in principle to predict the evolution of nonlinear systems without a real-time step-by-step simulation. In other words, the information necessary to describe the behaviour of these systems is *incompressible*:[15] there are no simple equations, shortcuts or patterns[16] by which compress the information about the evolution of a nonlinear system. In information theory, the degree of incompressibility of an object is measured through the so-called *Kolmogorov complexity*, which defines how much computation is required in order to specify the object at issue.[17] An example is the following: suppose to have two strings of fifty numbers that must be somehow transmitted. In the first case, the string is the following:

S_1 = 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2

In such a case, the transmission of the string will not require fifty numbers, as there is a pattern that governs the sequence which allows us to *compress* the original information S_1 in a smaller string of bits like "'12' 25 times".

By contrast, the second string, S_2, is a random sequence, such as:

S_2 = 6 0 9 7 7 5 8 3 4 2 6 6 7 4 6 1 2 0 5 8 9 7 8 6 8 7 6 9 5 1 8 7 8 2 5 8 1 5 6 4 4 4 7 0 9 3 5 6 3 9

In this second case, there is no pattern governing the formation of the string and this absence forbids compression. Therefore, to transmit the string, we shall use the same number of bits of the original sequence, without possible shorter encodings.[18] While the first string is compressible, the second is not, and we can say that this last one shows a greater Kolmogorov complexity.

Let us now consider how these concepts enter the debate about emergence.

[15] See Chaitin 1975: 48 "[…] the information embodied in a random series of numbers cannot be "compressed", or reduced to a more compact form".
[16] See Dennett 1991.
[17] A similar idea is Wolfram's incompressible computation, see Wolfram 1985 and 2002.
[18] See Chaitin 1975: 48 "A series of numbers is random if the smallest algorithm capable of specifying it to a computer has about the same number of bits of information as the series itself".

3.2.2 *Nonlinearity and weak emergence*

In 1997, Mark Bedau wrote a paper identifying "underivability without simulation" as the mark of a particular kind of emergence he defined "weak emergence". Differently from metaphysical or "strong" emergence, which Bedau judges as scientifically irrelevant, weak emergence is metaphysically innocent, it is used in science – so it is scientifically relevant – and it is even consistent with materialism, representing a useful tool to describe reality without conflicting with a scientific informed common sense. Bedau's weak emergence is shown in cases where there are systems composed by a certain number of micro-level components (microstates) that produce some macro-level states (macrostates) through a certain micro-dynamics *D*. The microstates are the intrinsic states of the components, while the macrostates are the structural properties of the system, which develop from the former. When these macrostates cannot be derived without a real simulation, the system exhibits what Bedau calls "weak emergence".

This idea can be described in the following terms: "Macrostate P of [the system] S with [local] microdynamic D is weakly emergent *iff* P can be derived from D and S's external conditions *but only by simulation* [emphasis mine]".[19] The necessity of a step-by-step simulation corresponds to the absence of a compression or shortcut to derive the development of the system, and in fact, some years after, Bedau has slightly modified his definition of weak emergence explicitly focusing on the notion of incompressibility:

> Elsewhere I have characterized weak emergence as underivability without simulation. Here I shift terms slightly and replace derivations with explanations, and replace macro-states that are underivable except by simulation with macro-states that have only incompressible explanations.[20]

Now, underivability without simulation, as well as information incompressibility, are not features a system has because of observers' cognitive limits. On the contrary, they reflect a *real* feature of certain systems whose dynamic can be derived only by simulation because of nonlinearity – being the last intended as interdependence between the components.

However, this point renders Bedau's weak emergence *really* weak, because even if just in principle and through the appropriate simulation, systems exhibiting weak emergence

[19] Bedau 1997: 378. Notice that the necessity of empirical observation is essential in Assad and Packard too: "Weakly emergent behaviour is deductible in hindsight from the specification *after observing the behaviour* [emphasis mine]" (Assad & Packard 2008: 232).
[20] Bedau 2008: 444.

have a derivable evolution. Bedau's weak emergence is metaphysically innocent, therefore, because it does not require the existence of novel entities, properties and/or causal powers – the macrostates derive from the microstates – and because this emergence admits physical derivability. A question arising at this point is why still talking of emergence if this account is so different from the historically received ones (a question Bedau himself poses at the end of his 1997 paper).

Traditionally, British Emergentists considered the presence of new causal powers and the impossibility of providing a good theoretical prediction as the marks of emergent phenomena. Bedau's account of emergence, in contrast, does not attribute these two features to the cases he acknowledges as instances of weak emergence. Nonetheless, he points out that in more recent literature other two important hallmarks have been assessed as characteristic of emergence:

> (1) Emergent phenomena are somehow constituted by, and generated from, underlying processes.
>
> (2) Emergent phenomena are somehow autonomous from underlying processes.[21]

Considering these two hallmarks, weak emergence seems a proper example of emergence: the macrostates are generated from the underlying microstates – as stated by (1) – but, at the same time, their behaviour is autonomous at the macro-level – as stated by (2): the necessity of a simulation, in fact, descends from the impossibility of a derivation based on the microstates and the external conditions alone. To understand this point, it is worth to outline the example Bedau takes in consideration, namely Cellular Automata (CA).

A CA is a spatially and temporally discrete dynamical system. It is composed by a regular grid of cells holding a finite number of possible states, such as black or white (two states), 0, 1 and 2 (three states), *etc*. For each temporal step t_1, $t_2 \ldots t_n$, every cell synchronously updates its state as a function of neighbours' previous states. The laws governing these interactions are simple, deterministic, local rules.

The simplest possible CA is called ECA (*elementary cellular automaton*) and it is a one-dimensional model (Figure 3.2) in which cells have only two possible states, and the neighbours relevant to each cell are those in close proximity (range=1). Under these

[21] Bedau 1997: 376.

conditions, at every time step *t*, ECA rules take into account the previous state of three cells: the updating one, the cell on its left, and the one on its right. Graphically, CAs rule sets are represented as in Figure 3.3, where, for instance, the first rule on the right (the first block) means: "if cell *a* is black and its neighbours are black, *a* remains black"; the second one: "if cell *a* is black, the left neighbour is black, and the right neighbour is white, *a* becomes white", and so on and so forth.

In ECAs with range equal to 1, the number of possible rule sets is 256[22], and, for every set, the emergent pattern is obviously different[23]. Figure 3.4, for instance, illustrates the pattern produced by the aforementioned rule set 110.

Now, in Bedau's framework, the macrostates af the CA – its macro-level patterns – are emergent because they respect the two hallmarks individuated by the author: first, constitution and dependence upon the processes; second, the autonomy given by the impossibility of a deductive derivation.

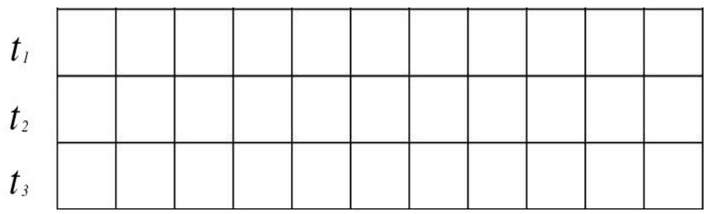

Figure 3.2 One-dimensional linear grid of cells. Every row corresponds to a discrete temporal step t_1, t_2, t_3 etc.

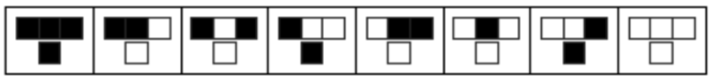

Figure 3.3. ECA rule set 110. The three top cells represent the eight possible combinations of neighbours' states, while the single bottom cell represents the updated cell. (Elementary Cellular Automata, Rule 110. From The Wolfram Atlas of Simple Program, http://atlas.wolfram.com/01/01/110/).

Bedau states as follows:

> […] the ontological and causal state of a cellular automaton macro structure is nothing more than the aggregation of the ontological and causal states of its micro constituents. At the same time, weak emergence exhibits a kind of macro autonomy because of the incompressibility of the micro-causal generative explanation of the macro structure.[24]

---

[22] Every cell can have two states (0 or 1), and every rule considers three cells, so we have eight possible configurations (23=8, namely 000; 001; 010; 011; 100; 101; 110; 111). Every rule set is therefore composed by eight single rules, as shown in Figure 5. The number of outcomes of every rule is again two (e.g. the updating cell can be 0 or 1), so the possible set of rules/combinations are 256 (28).

[23] See Berto & Tagliabue, 2017

[24] Bedau 2011: 97.

This is for unpredictability. However, Bedau emphasises another feature, namely that “the sciences of complexity are discovering simple, general macro-level patterns and laws involving weak emergent phenomena”.[25] Even from a nomological point of view, therefore, higher-level macro-phenomena seem partially autonomous because governed by distinct macroscopic natural laws.

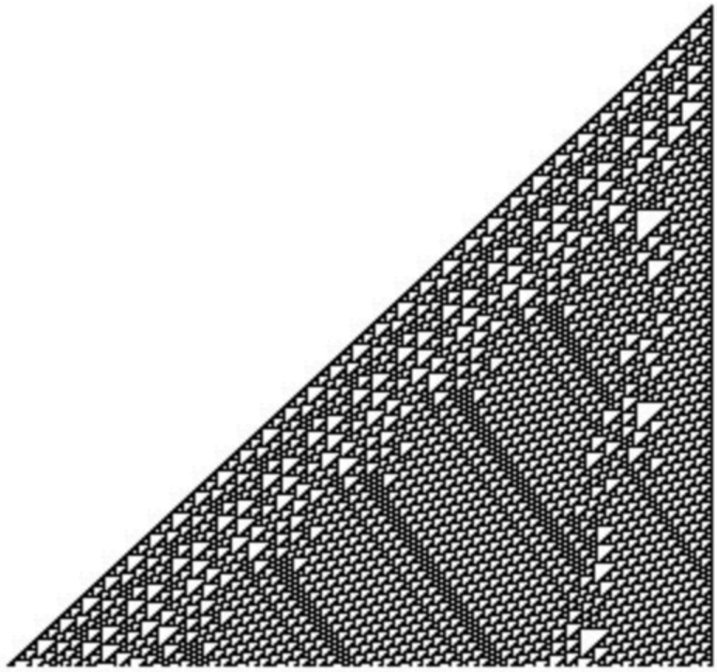

Figure 3.4 From the top to the bottom, the evolution of the CA following rule 110. This CA shows that from simple cells and rules the emerging pattern could be ordered and sophisticated. (Elementary Cellular Automata, Rule 110. From The Wolfram Atlas of Simple Program, http://atlas.wolfram.com/01/01/110/).

To sum up, in Bedau’s frame nonlinearity is responsible for the emergence of macrostates that are autonomous at the macro-level but are also constituted and generated by the lower-level microstates. Moreover, the powers of the macrostates are realised by those of the microstates, so the name “weak” for this kind of emergence seems appropriate because no further entities in addition to these macro-level patterns appear. The point that remains to clarify is whether this autonomy can be considered a sort of novelty, but the answer to this question is far from being straightforward, because it depends on what we are ready to accept as novel. Jessica Wilson, for instance, provides a negative answer to the issue because she is committed to a causal account of emergence and, consequently, to a causal account of novelty. We will focus on her view in the next paragraph. What can be anticipated, however, is that Wilson’s account of emergence takes as the mark of both weak and strong emergence the presence of new causal profiles or new causal powers. In her view, having novelty is having new causal powers. However, if we accept a more pluralistic criterion for novelty and consider relational properties and structures, or even qualitative properties, as genuine

[25] Bedau 1997: 395

novel features, the absence of new causal powers or activities will not directly involve the absence of novelty *tout court*.

In short, there are macroscopic laws that describe macroscopic phenomena without entering into the merits of their microscopic components (realizers). Also from a nomological point of view, therefore, the higher level macro-phenomena seem partially autonomous because they are governed by distinct and separate natural macroscopic laws. Summing up, according to Bedau, non-linearity is responsible for the genesis of macro-phenomena that are autonomous, even though they are constituted and generated by lower level micro-states. Bedau adds, moreover, that the powers of macro-states are realised by those of micro-states, so the name "weak" for this type of emergence seems appropriate because, traditionally, no emergence in the strong (or ontological/metaphysical) sense is attributed without the presence of new causal powers. The point that remains to be clarified is whether this autonomy can actually be considered a kind of novelty, but the answer to this question is far from simple because it depends on what we are ready to accept as novelty. Jessica Wilson, for example, thinks it makes sense to answer the question negatively, as it is engaged in a causal model of emergence and, consequently, in a causal model of novelty. I will focus on her point of view in the next paragraph, but what we can anticipate is that Wilson's emergence model recognises the presence of new causal profiles or new causal powers as an essential condition for emergence (be it weak or strong). According to Wilson, therefore, exhibiting novelty means possessing new causal powers. However, if a more pluralistic criterion is accepted also for novelty - and not only for emergence - relational properties and structures, as well as qualitative ones, can also be considered relevant in the attribution or otherwise of novelty.

3.2.3 *Nonlinearity and ontological emergence*

As we have mentioned, in Jessica Wilson's view nonlinearity is not a sufficient condition for emergence because "at least some cases of nonlinear complex systems – e.g. those associated with population growth or cellular automata[26] – clearly do not involve any additional fundamental forces/interactions, etc".[27] In Wilson's terms, several nonlinear systems are "physically acceptable" and for this very reason nonlinearity cannot be

[26] About this last example, see Bedau 1997.
[27] Wilson 2013: 207.

considered the mark of metaphysical emergence – i.e. the mark of a phenomenon that always implies the appearance of new causal powers or, at least, new causal profiles. However, Wilson's definition of physical acceptability – which corresponds to the exclusion of both novel features and emergent phenomena – is rather peculiar. She defines the physical as "fundamentally non-mental", committing herself to an account she defines "the physics-based NFM (no fundamental mentality) account":

> An entity or feature existing at a world *w* is physical if and only if (i) it is treated, approximately accurately, by current or future (in the limit of inquiry, ideal) versions of fundamental physics at *w*, and (ii) it is not fundamentally mental (that is, does not individually either possess or bestow mentality).[28]

This definition of the physical avoids the problem highlighted by Crane and Mellor and discussed in § 2.3, namely the "Hempel Dilemma" appearing when physicalists define the physical starting from what physics poses as physically acceptable. To summarise, in Joseph Baltimore's words:

> If the physicalist's account of the physical relies on current physics, then physicalism is implausible, for it seems highly unlikely that current physics is complete. If, on the other hand, the physicalist characterizes the physical in terms of future, ideal physics, then physicalism lacks determinate content, for it is unknown what entities future physics will end up positing.[29]

Despite not relying on physicists' statements and having therefore an advantage in respect to other accounts of the physical, the NFM view is still problematic. As for the first condition, it seems too strict, for fundamental physics (both current and, presumably, future) treats fundamental physical entities, but the physical domain seems to include non-fundamental physical entities as well (an atom, for instance, is a physical entity but it is not fundamental because it is composed by subatomic particles and held together by more fundamental powers). Adopting the first condition, therefore, may be inconvenient because it would exclude many non-fundamental but still physical entities, rendering Wilson's NFM account too strict. As far as the second condition is concerned, as suggested by Baltimore, it

[28] Wilson 2006: 74.
[29] Baltimore 2013: 14. See also Hempel 1969 and 1980, Crane & Mellor 1990 Crook & Gillett 2001 and Wilson 2006.

imposes a restriction on what future physics will discover and posit, building "too much metaphysics into the notion of the physical".[30] Imposing restrictions to future physics, however, is what is made by other authors too. David Papineau, for instance, highlights that the problem raised by the second horn of Hempel's dilemma – the one related to the fact that we don't know what a complete physics will posit – is just apparent, because rather than knowing what will be included in future physics, what is crucial for physicalist arguments "is to know what it [*future physics*] won't include".[31] Similarly to Wilson, thus, Papineau supposes that the physical will never be characterised as mental, sentient or intentional, and considers this last point historically motivated as well. Actually, as pointed out by Wilson too:

> […] physicalism is the descendant of materialism; and materialism is not only a fundationalist thesis but an anti-dualist one, in that mentality – typically understood in terms of the two traditional 'marks of the mental' – qualitative experience and intentionality - is supposed not to exist at the (relatively fundamental) foundations.[32]

Here, Wilson refers to what Seth Crook and Carl Gillett state about physicalism in 2001,[33] namely that accounts of physicalism just relying on physics cannot overcome Hempel's dilemma because physicalism should be viewed as a theory that rather than answering to scientific questions, should tackle philosophical and metaphysical ones, just as materialism did before being discredited by Positivists.[34] Crook and Gillett, therefore, suggest a switch from a scientifically based account of physicalism to a philosophical account that is more similar to materialism, and what should be noticed here is that the basic claim of materialism was precisely the idea that "mental entities are not amongst the ontologically basic entities".[35]

Now, while it might seem intuitively true that fundamental physical entities such as subatomic particles do not instantiate mental properties or intentionality (even if panpsychism affirms something different), defining in general terms the property of *being*

[30] Baltimore 2013: 15.
[31] Papineau 2001: 12.
[32] Wilson 2006: 69.
[33] Crook & Gillet 2001.
[34] See Crook & Gillet 2001: 347 "To summarize, our diagnosis is that the Positivist's wish to eschew metaphysics, and its concepts, led them, and the many philosophers who have recently followed their lead, to use the concepts of the physics in their formulations of physicalism to ultimately disastrous effect".
[35] *Ivi*: 348.

physical as the property of *being fundamentally non-mental* is a thorny, more serious issue, which can raise problems and criticism, as we have seen. We do not know what future physics will be like, we do not know exactly what the mental is (this is one of the reasons for the longevity and complexity of the debate on dualism and physicalism) and, consequently, to define the physical as essentially non-mental seems an arbitrary decision, aimed at protecting a metaphysical position taken as true a priori.

Now, to return to the question of non-linearity, we must take into account that, in the past, more traditional models of physicality and physical acceptability have been formulated and they too seem to suggest that nonlinearity does not necessarily imply ontological emergence. As Wilson has already pointed out, many complex dynamic systems are fully deterministic, where determinism can be defined by exploiting Robert Bishop's notion of *unique evolution*: a system is deterministic when each of its states is always followed (and preceded) by the same history of state transitions.[36] In other words, an S system is deterministic if it undergoes the same evolution every time it is restored to its original condition. Determinism, however, does not mean predictability: chaotic behaviour, which is a characteristic of many non-linear deterministic dynamic systems, cannot be predicted with precision due to the sensitivity of these systems to their initial conditions and our inability to specify them with sufficient accuracy.[37] Therefore, a form of epistemological emergence is always implicit in this type of systems, whose behaviour cannot be explained by the so-called "micro-reductive" explanations, i.e. those explanations that only take into account the microscopic level.[38]

What remains to be understood is whether non-linearity also implies some sort of ontological emergence and whether determinism is sufficient to exclude novelty and/or ontological emergence. As for the first question, the answer is negative for a number of authors. We have seen that for Wilson many nonlinear systems are physically acceptable, and therefore non-linearity, in her view, should not be considered as a sufficient condition for ontological emergence (incidentally, however, according to Wilson linearity is, on the contrary, a sufficient condition to *exclude* ontological emergence.)[39] Another author who suggests that nonlinearity does not introduce anything really new into reality is Stephen Kellert who, talking about chaotic systems, clarifies the importance of understanding that chaos theory argues against the universal applicability of the micro reductionist method, but

---

[36] Vescovo 2003: 170.
[37] See Kellert 1993: 62 "Chaos theory presents us with examples of systems that are described by differential equations, follow a unique evolutionary path, have certain values, yet are not predictable".
[38] On this, see Silberstein & McGeever 1999.
[39] Wilson 2013: 211.

not against the validity of the philosophical doctrine of reductionism.[40] For Kellert, chaos theory does not introduce any new postulate about the physical world, and in fact the models describing chaotic systems are built in a 'classical' world, using Newtonian equations.[41] For Kellert, therefore, non-linearity implies at most an epistemological novelty, because micro reductionism cannot explain chaotic behaviour. However, the metaphysical vision implied by reductionism is not compromised, because non-linearity does not guarantee the appearance of truly new entities or properties in the natural world.

As far as the second question is concerned - whether determinism is sufficient to exclude novelty and ontological emergence - it must be said that, even though they are often deterministic and "physically acceptable", non-linear dynamic systems manifest some interesting features that some scholars believe can reasonably be considered ontologically emergent. Silberstein and McGeever (1999), for example, note that complex systems manifest relational properties belonging to the system as a whole, but not to its fundamental parts. This circumstance suggests a kind of relational holism able to explain the fact that complex systems, regardless of their microphysical composition and the surrounding environment, show universal behaviour patterns that are robust with respect to internal or external changes, fluctuations and perturbations. This characteristic, also called *universality* o*r dynamic autonomy*,[42] requires an explanation[43] that can refer to emergence. The two authors believe, in fact, that one way to explain this dynamic autonomy – as well as the multiple realizability it implies – may be to assume the existence of ontologically emergent properties that *constrain* the properties and behaviour of the parts through a kind of downward determination. Without postulating the existence of these properties, the reason why many different systems, composed of different constituents, behave in the same way at the macro level remains a mystery, a miracle or a highly improbable "brute fact".[44]

One of the key features of complex dynamic systems, therefore, is to present, at the macro level, universal or *robust* behavioural patterns that remain constant despite microscopic fluctuations and changes. These patterns are called *attractors* and represent those phase spaces (i.e. the possible states of the system) to which the system spontaneously tends. Silberstein and McGeever recognise non-linearity as the cause of universality and consider this trait to be the reason for the failure of supervenience. According to the authors, non-

[40] Kellert 1993: 90.
[41] Ivi: 41.
[42] This definition is due to William Wimsatt. See Wimsatt 1981 and 1994.
[43] Silberstein & McGeever 1999: 196.
[44] Ibidem.

linear relationships are involved in what Paul Teller calls "relational holism", i.e. the condition that derives from the failure of supervenience for the relational properties that are not determined by the individual properties of the single *relata*.[45] For Teller, in other words, there are relational properties that do not supervene on the non-relational properties of their physical base. Now, if the mereological supervenience involves an exclusively vertical determinism, Teller's relational holism is a more complex model in which the relations of determination are vertical, horizontal and transversal. In non-linear systems, therefore, the components are related or mutually dependent,[46] which means that the network of dependency relationships in which they are captured constrains their individual characteristics and powers (and these constraints work from top to bottom, i.e. they are limitations that systemic patterns place on low-level components). The following could therefore be hypothesized.

First of all, these relational structures are universal structures that appear in various contexts emerging from lower level bases constituted microphysically in a heterogeneous way (which is similar to what Silberstein and McGeever suggested): this universality and robustness would also guarantee their autonomy on an ontological level. Secondly, these universal structures seem to all intents and purposes to be *determinative structures* whose causal profiles are (i) *new*, because they are different from those instantiated by their components, and (ii) *autonomous*, because they are resistant to changes and disturbances in their (robust and universal) emergence bases. These new causal profiles, as we shall see and as we have already partly mentioned, are represented by coercive and binding determinative powers, able to limit the possible states of the system's component parts. For this reason, it may be sensible to define these powers as "descending", i.e. able to determine not only other entities placed on their same level of organisation (vertical determination), but also entities placed on a simpler level of organisation (top-down vertical determination).[47]

Summing up, we have said that non-linearity implies a certain epistemological novelty understood as unpredictability, so it is certainly an indication of a weak or epistemic emergence. It remains to be determined, at this point, whether it is also an index of ontological emergence. When adopting a model of ontological novelty defined by the

[45] Silberstein & McGeever 1999: 197.
[46] It is no coincidence that in the science of complexity - as well as in the theory of information and probability - one of the ways of measuring complexity is based on the notion of mutual information. Mutual information measures the amount of information that one variable of a system can transmit on the others and is precisely a measure of the mutual dependence (i.e. the correlation) between the variables.
[47] I will return to this point, which is central to my discussion of emergence, in the next chapter.

presence of new individual causal powers, non-linearity cannot be considered to be a characteristic that directly implies ontological emergence. Wilson, who employs this model, consequently concludes that non-linearity is not sufficient for emergence to occur. However, if we assume a less strict model of novelty and consider ontologically relevant, in addition to individual causal powers, other properties, such as relational, structural, systemic or spatial properties, non-linearity may represent one of the conditions of ontological emergence, since it is able to produce *robust* and *universal* structures that exert some form of causal determination.

If non-linearity depends on the interdependence of the parts and their ability to determine each other, then the relational structures that emerge from these interactions will *bind* the parts, so that the behaviour of the latter, when involved in complex systems, will change with respect to them being isolated or inserted in linear systems. The determination we are talking about, therefore, could be defined as *binding* and it should be considered as an authentic form of causal effectiveness, even if different from the direct effectiveness exhibited by individual powers or dispositions. The difference between these two types of determinativity, one binding and indirect, the other active and direct, in my opinion is central to unraveling some theoretical issues that weigh on the debate on emergence, so this problem will be the subject of the next chapter. The next paragraph, however, will already allow us to address this issue because, as we shall see, in order to understand Samuel Alexander's emergentist vision and his interpretation of ontological novelty in terms of fundamentality, it is necessary to consider the nature of emergent causal powers and their difference from causal powers traditionally understood.

3.3 *Fundamentality*

Historically, emergence has been described through different formulations of the notion of fundamentality. In this paragraph, we first focus on the notions of *bruteness* and *brute fact*, which trace back to Samuel Alexander and British Emergentism, and were afterwards recovered by authors such as McLaughlin, Bedau and Wilson. Then, we will move on to the more contemporary notion of *fundamental novelty*, which involves the notion of *grounding* and where fundamentality is intended as ungroundedness, namely as lack of foundation and independence or *ontological priority*.

3.3.1 *Brute facts*

One of the features conventionally associated with emergence is bruteness, a notion descending from the idea of *brute fact* introduced in the debate by Samuel Alexander's *Space, Time and Deity*. Many authors – also among the British Emergentists themselves – more or less explicitly quote Alexander reporting his alleged definition of emergent qualities as 'brute facts to be accepted with natural piety'. In the modern debate, Brian McLaghlin established the correspondence between bruteness and fundamentality in his 1992 paper, *The Rise and Fall of British Emergentism*, which exerted a huge influence in the next debate about emergence. Discussing about another British Emergentist, C.D. Broad, McLaughlin states as follows:

> Emergent trans-ordinal laws are, Broad holds, brute nomological facts that "cannot be explained" […]. They must, he says, "simply be swallowed whole with that philosophic jam which Professor Alexander calls 'natural piety" (p. 55). *They are fundamental, nonderivative laws* [emphasis mine].[48]

McLaughlin, who restated this opinion some years later,[49] just focused on nomological trans-ordinal facts, namely on emergent *laws*, and considered them brute because fundamental. Another influential scholar, Mark Bedau, extends this idea to ontological emergent powers considering them "brute natural phenomena":[50]

> By definition, such causal powers [*strong emergence causal powers*] cannot be explained in terms of the aggregation of the micro-level potentialities; they are primitive or "brute" natural powers that arise inexplicably with the existence of certain macro-level entities.[51]

In more recent times, eventually, Jessica Wilson declares that her "conception of higher-level efficacy […] is one that is intended to be incompatible with physicalism, and is characteristic of British Emergentism".[52] To provide a further explanation of her statement, she quotes Lloyd Morgan who mentions again Samuel Alexander, stating that: "[…] what

[48] McLaughlin in Bedau and Humphreys 2008: 43.
[49] Ibidem.
[50] Bedau 2002: 11.
[51] Ibidem.
[52] Wilson forthcoming: 61.

emerges at any given level affords an instance of what I speak of as a new kind of relatedness of which there are no instances at lower levels […]. This we must accept 'with natural piety' as Mr. Alexander puts it".[53] The kind of novelty Wilson attributes to British Emergentists, therefore, is a fundamental novelty that is "incompatible with physicalism". By saying that, she is implying that a feature exhibiting this kind of novelty should be *accepted* (with natural piety), rather than *explained*, because this feature would be fundamental – the characterisations given by McLaughlin and Bedau to brute facts.

Now, it is worth making some considerations about this view because the expression "natural piety", as well as the notion of "brute fact", is usually traced back to Samuel Alexander but he did not consider the presence of emergent qualities as incompatible with physicalism. Alexander did not really say that emergents *are* brute facts, he says that their existence should be accepted "under the compulsion of brute empirical fact",[54] which means that the empirical evidence we have provides compelling proofs of the existence of these new qualities, which should be consequently acknowledged as existent, and studied with the same rigour and attention of other physical phenomena.[55] Here the complete passage:

> The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I should prefer to say in less harsh terms, to be accepted with the 'natural piety' of the investigator. [56]

Samuel Alexander, in this passage, uses the expression "natural piety", which became of common use in the debate about emergent features, but did not coin it. "Natural piety" comes from a poem written by William Wordsworth in 1802, *My Heart Leaps Up When I Behold*:

> My heart leaps up when I behold
> A rainbow in the sky:
> So was it when my life began;
> So is it now I am a man;
> So be it when I shall grow old,
> Or let me die!

[53] Lloyd Morgan in Wilson forthcoming: 62. It should be noticed that Wilson presents British Emergentists' views as a sort of system of theories, while it is clear that the differences among these scholars were relevant, especially as far as higher-level novelty was concerned.
[54] Alexander 1920: 47.
[55] About this, see Gillett 2006 and Symons 2018.
[56] Alexander 1920: 47.

The Child is father of the Man;
And I could wish my days to be
Bound each to each by natural piety.[57]

The expression "natural piety" is here used to indicate the genuine and spontaneous reverence for nature that is so common in the child, but that in the grown, cultivated man becomes weaker and weaker because of education. Scientific knowledge, but also other kinds of culture, such as the religious view of the world, "overwrite" new meanings on natural phenomena and the childish ability to immediately feel and see their pure nature is lost among all these complex interpretations. Coming back to Alexander's use of the expression, accepting something with natural piety – or under the compulsion of brute facts – should be understood as a pre-theoretical acknowledgement of the existence of natural phenomena opposed to a prescriptive metaphysical stance for which something is declared as *not really existent* – even if there is empirical evidence of it – because of some (physicalist, idealist or religious) reductionist judgements.

Now, Alexander, as we said, considered emergent qualities as new genuine properties of the world, but differently from what is usually handed down by interpreters, this opinion was not meant to weaken or question a naturalistic – we would say "physicalist", now – view of the world. As pointed out by Carl Gillett, who provided an insightful analysis of Samuel Alexander's emergentism,[58] emergent qualities are described by this scholar as *new* qualities that are arising from configurations of lower-level material components without being really different from them. This statement seems contradictory, because it is not clear how it is possible for something to be at the same time identical to something else and still distinct and autonomous from it. This contradiction has been acknowledged since McLaughlin 1992 paper,[59] where he states: "I am hesistant in my interpretation of Alexander, since, to be frank, I find apparently conflicting passages in his texts and I am uncertain how to resolve the apparent conflicts". The conflict at issue, here, is that produced by Alexander's complex formulation of physicalism, in which "higher level properties are realised by combinations of lower level properties and relations, but where these higher level properties are nonetheless causally efficacious".[60]

---

[57] Wordsworth in Roe 1992: 72
[58] See Gillett 2006.
[59] See McLaughlin 1992: "I am hesistant in my interpretation of Alexander, since, to be frank, I find apparently conflicting passages in his texts and I am uncertain how to resolve the apparent conflicts".
[60] Gillett 2006: 262-3.

The present discussion is the same as the one offered in the first chapter, in reference to Carl Gillett's notion of Strong Emergence explicitly recalled by Alexander.[61] For these two authors, emergent properties are both realised and causally efficacious (or new, as it is written in some passages) and according to Gillett, by "causally efficacious properties", Alexander means properties whose instantiation in an individual actually determines its causal powers.[62] In Alexander, therefore, emergent qualities are realised qualities that nevertheless play an effective role at an ontological level (they are not epiphenomenal) which consists in determining the type of causal contribution of their microphysical realizers.

This double structure of realization/determination is possible thanks to a particular conception of the microphysical realizers which, according to Alexander, are not homogeneous in their causal contribution, but rather "conditioned". As reconstructed by Gillett, Alexander denies that the instantiations of the microphysical properties are causally contributing in the same way each time. He also adds that, although the realized properties do not contribute causally as the fundamental properties, they are nevertheless efficacious in determining what will be the causal contributions of their microconstituents.[63] Therefore, on the one hand, in accordance with physicalism, Alexander states that microphysical properties are the only properties able to contribute causally through original individual powers. On the other hand, in addition to the fundamental microphysical properties that make this kind of causal contribution, there are other properties that have a determinative efficacy, since, as we have seen, they are able to define what will be the causal contributions of the microconstituents to be or not be effectively exercised.

So, if a property is usually considered efficacious just when directly contributing a causal power, in Alexander's account, by contrast, a property is causally efficacious also when it *determines* or *constrains* the contribution of a causal power, and this wider interpretation of "mediated" causal efficacy is what allows Alexander (but also Gillett[64]) to consider emergence as compatible with realisation and physicalism (but not with the closure of the physical). We will return on these notions in § 3.6. What was relevant here, however, was to disentangle Alexander's use of the term "brute" from the concept of ontological fundamentality. In Alexander's view, emergent qualities are realised qualities, therefore they

[61] Gillett's account of emergence is "just a reformulation of Alexander's account" (personal communication).
[62] Gillett 2006: 264.
[63] Gillett 2006: 274.
[64] Gillett's account of emergence is "just a reformulation of Alexander's account" (personal communication).

cannot be considered metaphysically fundamental even if they must be accepted in our ontology given their peculiar metaphysical role in determining their own realisers.

A merely empirical connotation of the notion of bruteness, however, is far from being generally accepted in contemporary literature. As pointed out by a certain number of scholars,[65] brute facts are facts for which there is no explanation, i.e. facts for which the Principle of Sufficient Reason fails. However, this failure may have a twofold characterisation that has been outlined by Eric Barnes. There are *epistemologically brute facts,*[66] when facts cannot be explained because an explanation is not available *yet*; and there are *ontologically brute facts*, when facts cannot be explained *in principle* because they are primitive, fundamental states of affairs that are just what they are, without further reason or cause: "facts with no explanatory basis beyond themselves".[67] It seems that Alexander's use of the term "brute" refers to empirical bruteness, and in fact both Alexander and Broad[68] state that is possible that in the future, thanks to new scientific discoveries, these unexplicable natural phenomena will be perfectly explicable, losing their status of emergents.[69] In many contemporary authors, moreover, the novelty emergence involves is considered of an epistemological kind, because produced by epistemological irreducibility. With the development of the debate about emergence, however, ontological bruteness came into the picture too, and the notion of novelty sometimes acquires the meaning of fundamental ontological novelty, the kind of novelty on which the next paragraph focuses.

3.3.2 *Ontological priority*

Assuming the existence of fundamental phenomena implies tracing distinctions in reality among the fundamental and the non-fundamental, as well as opening the question about how the former and the latter are related. As we saw in §2.3.1, many philosophers use a theological metaphor to describe the main feature of the fundamental – or its consequence,

---

[65] See Barnes 1994, McKaughan 2013, Symons 2018.
[66] Barnes 1994: 62.
[67] *Ibidem*. See also Cameron 2008: "A fact *f* is *brute* iff there is no fact *g* such that *f* holds in virtue of *g*; *f* is *derivative* otherwise".
[68] See Broad 1923: 55 "It was held that the characteristic differences between the behavior of Oxygen and Hydrogen are due in no way to differences of structure or components, but must simply be accepted as ultimate facts. This first alternative can hardly be counted as one way of explaining differences of behavior, since it consists in holding that there are certain differences which cannot be explained, even in part, but must simply be swallowed whole with that philosophic jam which Professor Alexander calls "natural piety". It is worthwhile to remark that we could never be logically compelled to hold this view [...] Nevertheless, it is perfectly possible that [...] there are certain ultimate differences in the material world which must just be accepted as brute facts.".
[69] About this, see Symons 2018.

at least. The metaphor, which was introduced by Saul Kripke in *Naming and Necessity*,[70] consists in identifying the fundamental with the answer to the following question: "what does God need to create to have the world we have?".[71] From Elisabeth Barnes:[72] "The fundamental entities are all and only those entities which God needs to create in order to make the world how it is". From Wilson: "The fundamental is, well, fundamental: entities in a fundamental base play a role analogous to axioms in a theory – they are basic, they are 'all God had to do, or create'".[73] From William Seager:[74] "once God created the physical world, set the physical laws and the arrangement of the fundamental physical entities in the world, there was nothing left to do about the non-fundamental things". In simple words: "[…] 'all God had to do' was to create the primarily real",[75] namely the fundamental. Once done that, everything else will be *freely* obtained.

Now, the reason why the fundamental is enough to have all what there is in the world – the reason why everything obtains once given the fundamental – is that the non-fundamental is supposed to reduce to, be determined by, depend upon, or be "grounded" in the fundamental.

In recent years, this last notion of *grounding* has been viewed as a concept able to unify all these different metaphysical relations of dependence, determination and reduction, as well as to shed light upon fundamentality and derivability. Jonathan Schaffer, for instance, writes:

> The notion of reduction is intended to be an ontological relation, expressing dependence between entities. […] As a relation of dependence, the intended notion of reduction may be glossed in terms of grounding. What reduces is grounded in, based on, existent in virtue of, and nothing over and above, what it reduces to. What does not reduce is basic, fundamental, and brute.[76]

For Schaffer, therefore, the fundamental is what cannot be reduced because *ungrounded*, i.e. primitive, ontologically prior, or brute.

---

[70] See Kripke 1972: 153 et seq.
[71] Actually, the question Kripke poses in his book is less wide: "Suppose we imagine God creating the world; what does He need to do to make the identity of heat and molecular motion obtain?". The type of reasoning, however, was later taken up in more general terms.
[72] Barnes 2012: 826.
[73] Wilson 2014: 540.
[74] Seager 2014: 146.
[75] Schaffer 2004: 100.
[76] Schaffer, 2008: 83.

Jessica Wilson expresses the idea in similar terms:

> It has recently been suggested that a distinctive relation – call it ('big-G') 'Grounding' – is at issue in contexts in which some entities, propositions or facts are claimed to 'metaphysically depend on' (in a constitutive rather than causal sense), 'hold in virtue of', be 'nothing over and above', or be 'grounded in' some others.[77]

The notion of Grounding, which originally traces back to Brentano, has been recovered and significantly debated in contemporary analytic metaphysics. Its importance is due to the fact that, on the one side, it seems to unify different important relations of metaphysical dependence – those called by Wilson "'small-g' grounding relations" – and, on the other side, it highlights their role in shaping the concepts of fundamentality and derivability. For these very reasons, however, while some authors like Schaffer[78] and Gideon Rosen[79] adopted it, many others, like Bennett[80] and Wilson[81] exhibit scepticism about its usefulness and consistency. The main problem about Grounding is that it is not clear whether there really *is* a unitary "'big-G' Grounding", for this word may be just an umbrella term including those "small-g"[82] metaphysical grounding relations listed by Wilson[83] and generally accepted in literature. Schaffer, hoping in a "revival of a more traditional Aristotelian view, on which metaphysics is about what grounds what [*and not only about what there is, like Quine recommended*]",[84] claims Grounding to be "the primitive structuring conception of metaphysics",[85] but a sceptic such as Wilson highlights the metaphysical underdetermination of Grounding:

> […] Grounding cannot do the work proponents of Grounding want it to do. For Grounding, like supervenience, is too coarse-grained to characterize appropriately metaphysical dependence on its own […]. Doing the job requires appeal to the 'small-

[77] Wilson 2014: 535
[78] See also Schaffer 2009.
[79] Rosen 2010
[80] Bennett 2017.
[81] Wilson 2014.
[82] The expressions "big-G Grounding" and "small-g grounding relations" have been firstly used by Wilson 2014.
[83] Regarding the debate concerning the legitimate use of the notion of Grounding see Wilson 2014, Bliss and Trogdon, 2016: § 1. *Is grounding unitary?* And § 8. *Skepticism about grounding*
[84] Schaffer 2009: 364
[85] Ibidem.

> g' grounding relations that have been traditionally appealed to in investigations into metaphysical dependence.[86]

Now, there is a complex debate[87] about the legitimacy, definition, and usefulness of the notion of Grounding and outlining this debate is beyond the scope of my work; however, it can be said that what Grounding is supposed to show is what entities are *metaphysically prior* to others, and in this framework fundamentality is considered the bottom of this layered view of reality hierarchically ordered by metaphysical priority.

Metaphysical priority traces back to Aristotle's *Metaphysics Delta*, 11, 1019a1, where he defines "priority in nature or in substance" in the following terms:

> Some things are called prior and posterior in this way, while others are called so in nature and in substance [κατὰ φύσιν καὶ οὐσίαν], those for which it is possible to be without other things.

As pointed out by Peramatzis, the meaning of εἶναι, here, can be twofold. It can mean "to exist" but also "to be what something is". In the first case, the priority is *existential*: "*A* is ontologically prior to *B* if and only if *A can exist* without *B existing* but not the other way around".[88] Otherwise, εἶναι can mean "to be what something essentially is", and in this second case "*A* is ontologically prior to *B* if and only if *A* can be what it is independently of *B* being what it is, while the converse is not the case".[89] While many authors interpreted Aristotelian priority in nature as mere existential priority, Peramatzis states that there are good conceptual and exegetical reasons to leave the interpretation open to both the existential and "essentialist" views.[90]

Beyond the debate about the Aristotelian interpretation of priority in nature, however, a deep and articulated analysis of the notion of ontological priority has been subsequently developed and the "essentialist" view attributed by Peramatzis to Aristotle is now particularly relevant. In recent times, Kit Fine analysed the notion of ontological dependence and observes that an existential and modal interpretation of ontological priority – "One thing

[86] Wilson 2014: 542
[87] About this see Correia & Schnieder 2012.
[88] Peramatzis 2011: 205.
[89] *Ivi*: 205.
[90] See Peramatzis 2011. Peramatzis' interpretation of priority in Aristotle is similar to Jonathan Beere's one. See Beere 2010: 293 et seq. Other authors who propose a mere existential interpretation are, by converse, Clearly 1988 and Panayides 1999.

x will depend upon another *y* just in case it is necessary that *y* exists if *x* exists”[91] – is “too loose”, because it does not really explain the reason why there is a dependence relation among *x* and *y.* In Fine’s opinion, the modal/existential account should be revisited as follows. On the one hand, the modal force of the dependence relation between *x* and *y* should be tied to the nature of the entities involved. It is *in the nature* of the dependent entity *x* to exist only when *y* is existing: it is *essential* to the identity of *x* to have this property. On the other hand, the focus on the of existence of *x* due to the existence of *y* depends upon a common assumption that Fine considers controversial, namely that the being of an entity should be identified with its existence. This assumption raises two problems: “In one respect, existence is too weak; for there is more to what an object is than its mere existence. In another respect, existence is too strong; for what an object is, its nature, need not include existence as a part”.[92] Now, as far as these problems are concerned, the notion of essence is useful again. Rather than identifying the being of *x* with its existence, it seems rightful to identify it with a collection of essential properties, *constitutive* of the nature of *x*. In this frame, it may be right to “[…] take *x* to depend upon *y* […] if *y* is a constituent of an essential property of *x*”,[93] or, from a definitional point of view, “if it [*y*] is ineliminably involved in all of its [of *x*] definitions”. This means that the being or essence of an entity corresponds to its “collection of essential properties”,[94] or “the collection of propositions that are true in virtue of its identity”.[95] Ontological priority, in this view, corresponds to the state of an object which grounds the existence and/or the essence of another object, which is ontological dependent on – and consequently less fundamental than – the former. In fact, as pointed out by Ross Cameron, ontological dependence is the converse of ontological priority. An entity *x* (such as a composite object) ontologically depends upon another entity *y* (e.g. one of its parts), and then *y*, as a consequence, is ontologically prior to *x*. The direction of these relationships, however, can be twofold: on the one side, the composite whole can be considered as dependent on the parts; on the other side, the parts can be considered as dependent upon the composite object. Historically, these two interpretations can be defined pluralism (or atomism) and monism, respectively:

[91] Fine 1995: 270.
[92] *Ivi*: 274.
[93] *Ivi*: 275.
[94] Ibidem.
[95] Ibidem.

> The *monist* view that what is basic is the entire universe and it is the *cosmos* that is ontologically prior and independent, while its parts are derivative and dependent, was held by the likes of Parmenides, Plato, Plotinus, Proclus, Spinoza, Hegel, Lotze and Bradley. It was then set aside in favour of the idea […] that what is truly fundamental are elementary entities more or less fitting the description of Democritus' atoms.[96]

While for monism – recently revitalised by Jonathan Schaffer – the whole is prior to the parts, "with metaphysical explanation dangling downward from the One"[97], for pluralism the parts are prior to the whole "with metaphysical explanation snaking upward from the many".[98]

Now, we saw that the question *what is prior to what* is equivalent to the question *what grounds what*, which is another version of the problem *what is fundamental*. In the framework outlined here, therefore, the fundamental can be considered as the *ungrounded*, namely the ontologically prior class of entities that (i) do not depend on anything else for their existence and essence and (ii) constitute the basis for everything else. This dependence, in literature, is unanimously defined as irreflexive, asymmetric and transitive. The question now is how does emergence fit in this frame.

As we extensively said, the most relevant features attributed to emergence are those of irreducibility – both ontological and epistemological – and novelty despite dependence (or even despite realisation, such as in Alexander and Gillett). Emergent phenomena, in other words, show novel fundamental features despite their compositional and causal dependence upon a lower-level, more elementary, and entirely physical base. Recently, Elisabeth Barnes clearly formulated the idea that emergent entities are "those which are fundamental but not independent". Let us outline her view.

Barnes considers both fundamentality and dependence as *metaphysical primitive*, so she does not define them, but she outlines some brief analyses that suggest as follows. On the one hand, fundamentality is described exploiting the theological metaphor and in correlation with mereological composition: "if God decides that she wants a world with a single complex object composed of two mereological simples, she would simply have to create the two mereological simples". The fundamental, therefore, is what God has to create to have everything else, and it seems that the relationship at issue, here, is one of

[96] Morganti 2009: 271.
[97] Schaffer 2010: 31.
[98] Ibidem.

composition between the simplest fundamental entities God needs to actively create and the derivative, dependent entities obtained for free. On the other hand, dependence is described in two different ways. There is a weaker sense of dependence, such as when "a person is dependent on her parents for her existence: her parents cause her to exist, and had her parents not existed she wouldn't have existed". In this case dependence is correlated to essentiality of origins in that the dependee *causes* the dependent.[99] In addition to this narrow sense, there is a stronger ontological sense of dependence, which holds in cases of complex objects and their compositional parts. In this case "rather than merely being counterfactually dependent on the existence of something in its past, the object is dependent *at each moment of its existence* on the existence of something *which exists at that very time*". Rather than being caused by the dependee, in this case the dependent is *sustained* by the former. Now, while, in general, fundamental entities are ontologically *independent*, emergents are fundamental entities that are ontologically *dependent*. This means, in Barnes view, that despite being dependent, emergents would have been actively created by God because no fundamental entity is obtained for free from derivation from something else. This point is relevant because Barnes' idea of fundamentality does not allow for relative fundamentality but just for absolute fundamentality. In her view, fundamentality does not come in degree. There are just fundamental entities and non-fundamental ones, as well as there are no more or less derivative entities, but just derivative entities or non-derivative (i.e. fundamental) ones. Barnes states that her view avoids the problem of the levels metaphor, because rejecting relative fundamentality implies adopting an almost flat ontology merely composed by a fundamental level and a derivative one. Now, Barnes' distinction between the fundamental and the derivative level focus on two criteria: fundamentality and dependence. However, we saw that emergents are both fundamental and dependent, so the two categories cut across each other, and given that, entities could be neatly distributed into four distinct metaphysical categories:

(i) the fundamental independent entities;
(ii) the non-fundamental dependent entities;
(iii) the fundamental dependent entities (i.e. emergents);
(iv) the non-fundamental independent entities (such as numbers).

[99] Barnes states that "an entity x is dependent on entities $y_1 \ldots y_n$ just in case x is both caused and *sustained* by the y's" (Barnes 2012). Then she proposes the distinction between weak and strong dependence, so I conclude that causation is linked to essentiality of origins, while sustainance to composition.

Nonetheless, the levels acknowledged by Barnes are just two, and emergents simply belong to the first one together with the other fundamental phenomena, such as mereological atoms; conversely, everything else belongs to the derivative level. This is shown by the images above, which represent a hierarchical view of fundamentality that Barnes does not accept (above), and Barnes' own fundamentalist view of reality (below).

Now, the idea of a neat distinction between fundamental and derivative entities can be useful to conceptually clarify the debate about emergence from a meta-ontological point of view (aim that Barnes declares at the very beginning of her paper). This framework, however, is supposed to avoid the problems arising from the levels metaphor denying the existence of more than two well-defined and totally different and non-overlapping levels, but this view leaves open the problems that made the levels metaphors significant and that made it to be formulated in the first place. The coexistence between dependence and the exhibition of novel properties or behaviours, together with the admission that despite not clearly defined,

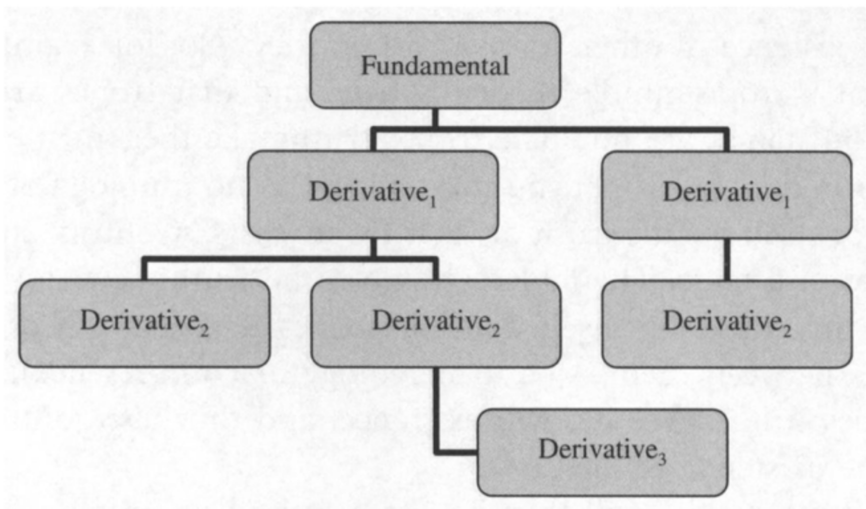

Figure 1: The hierarchical view of fundamentality

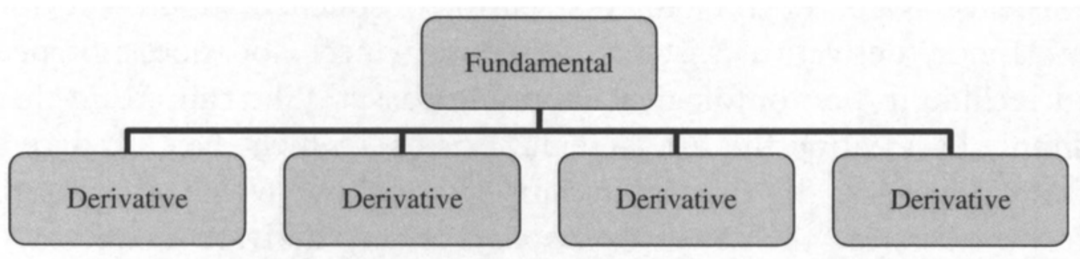

Figure 2: The fundamentalist view of fundamentality

Figure 3.5. The two view of fundamentality delineated by Barnes (2012)

emergence seems a common feature of the world, suggests that emergent fundamentality might be scattered throughout reality rather than localized at a particular level – the lowest, as suggested by atomism, or the highest as implied by monism. This point is relevant, as many scholars have pointed out that the levels metaphor is too strict and monolithic and "surely has to be articulated in less crude – and likely more piecemeal and contextual – terms […]".[100] Now, the necessity for a fundamental level able to b ase and limit the chains of dependence is perfectly reasonable: "there is a problem if metaphysical explanation never 'grounds out' at some fundamental level. While you can have a fundamental level and infinitely many things dependent on that level, you cannot have dependence all the way down".[101] The assumption of the existence of infinite downward dependence chains – the so-called *metaphysical infinitism* – leads to the problem of infinite regress, which is something it is wise trying to avoid. However, nothing prevents to assume that the bottom grounding basis of these dependence chains could be multiple and spread at each "level of organisation", rather than belonging to a single kind of object or to a particular *stratum* of reality. Revisiting the levels view, in conclusion, can be useful to make it compatible with emergence and its peculiar features, which seems naturally fitting in a layered view of reality, rather than in a too much monolithic hierarchical picture of the world.

3.3.3 *Levels*

The world in which we live appears to be an integrated whole, therefore the elaboration of an ideal, integrated knowledge about the world has been a permanent goal in philosophy and science[102]. Despite unity, however, reality exhibits diversity as well, so a view of the world including various kinds of entities naturally emerged during the history of philosophy. This view gives rise a further question, namely how these different kinds of object are related among each other. To answer this question, a hierarchical view of entities organised into different levels or layers of "being" has been formulated. As stated by Kim,

> […] the shared imagery evoked by levels talk is a picture, somewhat fuzzy and unarticulated, of the physical world neatly stratified into a structure of discrete levels, with a bottom level of basic particles – perhaps leptons and quarks, or whatever our best

[100] McKenzie in Gibb, Hendry and Lancaster 2019: 56.
[101] Cameron 2008: 3.
[102] See Oppenheim & Putnam 1958, for a review, see Cat 2017.

physics tells us are the fundamental constituents of matter – and the rest as forming a vertically ordered system of levels each resting on the one below and all ultimately resting on the base level of microparticles.[103]

This hierarchical view of reality implies that entities can be organised from the bottom to the top, from the simplest, most basic and fundamental ones to the most complex, structured and derivative ones. Although *prima facie* intuitive, conceptually neat and elegant, this picture is hardly understandable from a more fine-grained metaphysical point of view, however, because many details remain unclear. On the one hand, it seems intuitive to state that particles are simpler than atoms, that atoms are simpler than molecules, that molecules are simpler than cells, and so on and so forth. This classical list traces back to the reductive levels enumerated by Oppenheim and Putnam in 1958

6……………………. Social groups
5……………………. (Multicellular) living things
4……………………. Cells
3……………………. Molecules
2……………………. Atoms
1……………………. Elementary particles[104]

On the other hand, however, a huge number or questions arise when this view is considered in detail. Paraphrasing Kim[105] and Carl Craver,[106] adopting the levels picture requires to clarify (i) how levels can be clearly distinguished and identified; (ii) what makes a given level higher (or lower) than another one; (iii) how a phenomenon should be placed in a particular level rather than in another one; (iv) whether there is a bottom level; (v) which kind of relation orders the different levels; (vi) what kinds of entity (*asbtracta* or *concreta*? Types or tokens? Substances, activities or properties?) are the *relata* of the ordering relation; (vii) whether it is reasonable to expect that each object belongs to just one level; (viii) whether there is a fixed system of levels (for maybe it is just a matter of conventional classification); and so on and so forth.

[103] Kim 2010: 43.
[104] Oppenheim & Putnam 1958: 9.
[105] Kim 2002: 4.
[106] Craver 2015: 3.

Now, Oppenheim's and Putnam's schema of reductive levels obviously offer an extremely simplified version of the levels view which is based on science and how scientific knowledge is organised. A more sophisticated version is the one C. Lloyd Morgan had provided in *Emergent Evolution* in 1923. Here, Morgan depicts a simple synoptic schema for emergent evolution: at the bottom there is spacetime (Morgan follows Alexander), which "extends throughout all that is".[107] At the top there is Deity, "an emergent quality that characterises only certain persons at the highest and latest stage of evolution along a central line of advance".[108] Among spacetime and Deity, there is "a vast multitude of individual pyramids"[109] where atoms, molecules, plants, animals and human beings emerge from entities belonging to "lower grades". At each level, new kinds of *relatedness* appear, and emergent qualities are the expression of these new relational structures that provide integrated unity to the entities appearing at the levels at issue.

Despite leaving open several of the questions mentioned above, Morgan's description of ontological levels highlights their *nested* structure: "On this understanding we distinguish mind, life, and matter. Within each, of course, there are many emergent sub-orders of relatedness".[110] Morgan's synoptic pyramid, therefore, is the simple and schematic expression of a multitude of further pyramids, so that "intermediate levels can be interpolated without end".[111] The nested view of levels seems to be less strict than the discrete and idealised picture provided by Oppenheim and Putnam, and Morgan's acknowledgement of the complexity of the ontological structure of reality is coherent with the already mentioned remark made by McKenzie, namely that the levels metaphor might be intended as "piecemeal", rather than "monolithic" – a point highlighted by William Wimsatt[112] and Carl Craver[113] too. Let us explore their suggestions.

3.3.3.1 *William Wimsatt's view of levels*

William Wimsatt describes his view about the organisation of reality in a long and detailed paper published in 1994, *The Ontology of Complex Systems*. Here, the author points out that talk about levels usually suggests the idea of a neat division of nature into robust

[107] Morgan 1923: 10.
[108] Ibidem.
[109] *Ivi*: 11.
[110] Morgan 1923: 22.
[111] Kim 2002: 6.
[112] Wimsatt 1994.
[113] Craver 2015.

and monolithic layers of reality. Wimsatt does not mention Oppenheim and Putnam, but Popper and his theory of the three worlds instead. The criticism, however, is the same:

> These rough distinctions are of major importance – delimiting regions where different major concepts, theories, methodologies, and explanatory strategies dominate, but they are larger heterogeneous aggregates spanning multiple levels and including also other less well ordered structures rather than single individual levels of organization.[114]

This division into levels, in other words, is useful but inaccurate, because reality is far more complex than what is suggested by this hierarchical and neat compositional view of the world. Within any of the levels identified by Popper (and the same can be said about the levels described by Oppenheim and Putnam) several other intermediate levels can be acknowledged and even in the case in which a huge number of levels is to be accepted, difficulties would still arise in their precise mapping. The reason why this operation would be hardly accomplished is that levels, in Wimsatt's opinion, are more easily identifiable at lower levels of complexity, where the degrees of freedom are just a few and matter can aggregate in quite simple forms. At higher levels of complexity, conversely, matter acquires a great number of possibilities of organisation and interaction, therefore entities become less defined, more multi-dimensional, and "richer in their budget of properties".[115]

In one sentence: "There should be more ways of interacting with a spouse than with a quark!".[116]

To understand this view of levels, it is worth noticing that for Wimsatt the "primary working matter of the world" consists in networks of causal relationships. These networks became more and more complex over time and organised themselves in different ways and in larger patterns. Among these patterns there are those called by Wimsatt "levels of organisation", which are "local maxima of regularity and predictability in the phase space of alternative modes of organization of matter".[117] Among all the ways and modes in which matter can aggregate and organise itself, that is to say, some *forms* are more stable and

[114] Wimsatt 1994: 5.
[115] Wimsatt 1994: 17.
[116] *Ivi*: 17.
[117] *Ivi*: 11.

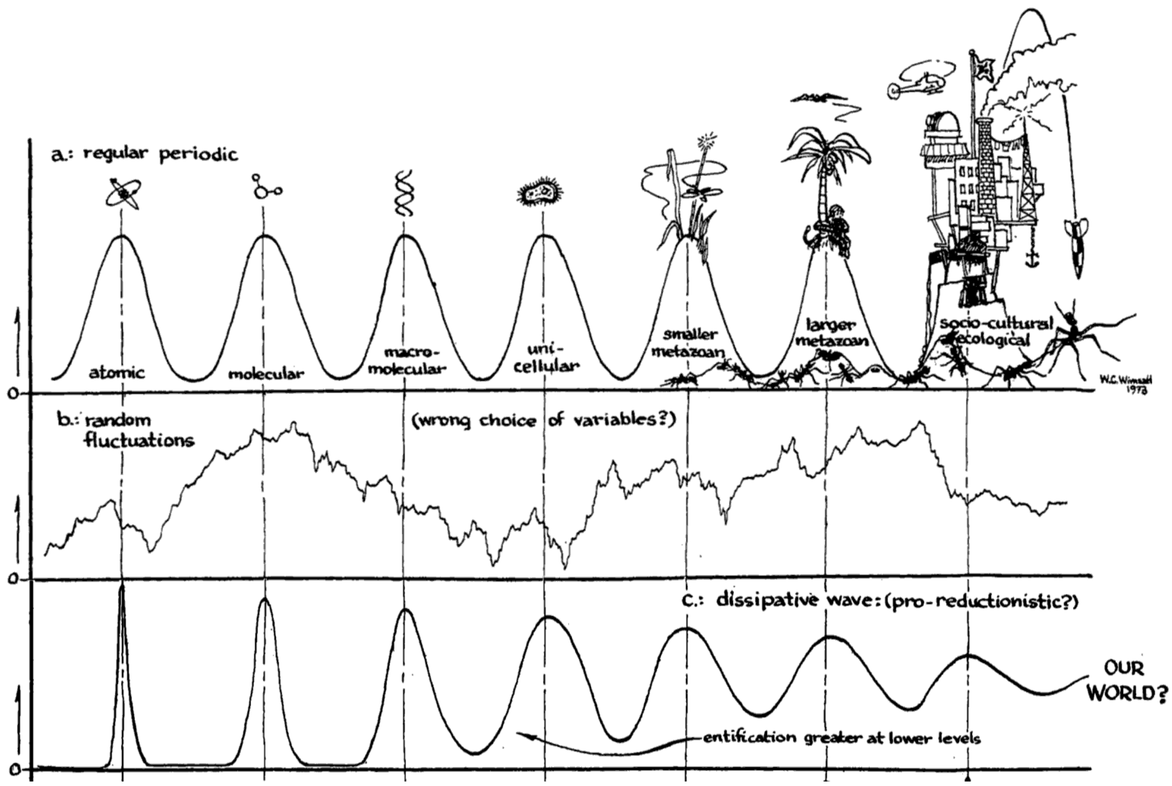

Figure 3.6 William WImsatt's view of levels as local maxima of regularity and predictability. Wimsatt 2007: 224-225.

predictable than others, a nd these forms correspond to the idea of levels of organisation (Fig. 3.1 represents this idea).

However, with the increase of complexity, levels become less and less well defined and neat, turning into something else, namely *perspectives* or *causal thickets*. In Wimsatt's view, therefore, reality is structured in various complex ways, and among these ways distinctive ontological structures (i.e. levels, perspectives, and causal thickets) can be found, spanning from the less complex, less context-dependent, and more regular – e.g. the level of the quark – to the more complex, more context-dependent, and less regular and predictable ones – e.g. the causal thicket to which the spouse belongs.

As for the ontological structures aforementioned, we already described levels. *Perspectives*, differently from levels, are less objective. They are "quasi-subjective (or at least observer-, technique- or technology-relative) cuts on the phenomena characteristic of a system, which needn't be bound to given levels". Perspectives are characterised by an *objective-subjective duality*, because they involve a set of variables that are salient to a particular observer, but do not offer a complete, objective description of the systems at issue. For instance, "anatomy, physiology, and genetics are different perspectives on an organism

[…]", but these perspectives cannot be related and described in compositional terms, because they are different, given the different sets of variables defining them. Similarly, the problems different perspectives aim to solve cannot be considered the same, because each perspective has paradigmatic problems that is able to solve without information coming from other perspectives.

There are further problems, however, that need two or more perspectives to be solved (for instance, the mind-body problem or the nature of language), so in certain cases, when there is a relevant "boundary ambiguity", perspectives degenerate into *causal thickets*, namely cases in which phenomena can neither be clearly placed into one level, nor defined by just one perspective. For example, "the neurophysiological, psychological, and social realms are mostly thickets, which are only occasionally well-ordered enough for local problems to be treated as perspectival or level-relative problems".[118] Causal thickets, however, are not a "waste-basket", ontologically speaking. On the contrary, they are the most common kind of systems, given the huge number of possible interactions available to complex entities and systems above the atomic or molecular level. In Wimsatt's view, therefore, rather than causal thickets, it is insulated levels of organisation and exhaustive perspectives to be rare in nature. Therefore, a relevant number of the questions raised above can now be answered or partially suspended: the ordering relations between levels of reality is composition, but just for levels of organisation, i.e. for the simplest and most elementary relational structures, and not for perspectives and causal thickets, which are not compositionally ordered. Similarly, outside the limited domain of the Wimsattian levels of organisation, it is possible neither to neatly assign a phenomenon to a particular level, nor to determine its position in relation to other entities because, as we saw, there is no clear and fixed compositional ordering among perspectives and causal thickets.

Wimsatt's main thesis, to conclude, is that reality is more complex than what the canonical levels metaphor suggests. Levels can only be detected at a low degree of complexity, where matter aggregates in simple forms. The vast majority of phenomena, instead, are organised in complex ways and this complexity cannot be accommodated in a metaphysical view based on such a rigid notion of absolute and monolithic layers of reality. If we still want to be committed to common structures of organisation characterising reality, we have to be more liberal and accept structures that are less frozen, more context-dependent, and more difficultly conceptualisable.

[118] Wimsatt 1994: 239.

3.3.3.2 *Levels of mechanism*

Carl Craver, together with Bill Bechtel, further developed this weakening of levels of being by formulating an account of "levels of mechanisms". In his 2015 paper, *Levels*, Carl Craver offers a clear analysis of the levels metaphor and describes his view of the problem. The author states that the metaphor usually "takes an apparently heterogeneous collection of objects and arranges them in space from bottom to top".[119] This picture, as we have already saw, makes some issues arise, but those on which Craver puts his focus are three: "*The Relata Question*: What kinds of item are being sorted into levels? *The Relations Question*: In virtue of what are two items at different levels? *The Placement Question*: In virtue of what are two items at the same level?".[120]

Now, starting from the second one, many kinds of relationship can be identified as relevant in the hierarchical ordering of levels. Levels can be levels of abstraction, analysis, and/or explanation – representing an epistemological and conceptual ordering of reality – but also, alternatively, levels of organisation, aggregation, composition, realisation, or causation – reflecting a more significant ontological commitment to the levels metaphor. Craver suggests that the most common ordering relation in the levels debate is the compositional, part-whole relationship.[121] However, classical mereology is not appropriate to characterise the ordering relation, for, as pointed out by the author, it is associated with a set of assumptions that are incompatible with real scientific decompositional practice. For Craver, the set of assumption characterising classical mereology consists of the following axioms:

(i) *reflexivity*;

(ii) *transitivity;*

(iii) *extensionality*;

(iv) *summation*.

Classical part-whole relationships, in other words, imply that (i) any object is part of itself, (ii) a part of a part of an object is a part of that object, (iii) between an object and its parts holds a relationship of complete determination, therefore "for objects to be identical, it suffices that they have all and only the same parts", and, finally, (iv) any pair of objects can be summed becoming a new object.

[119] Craver 2015: 1.
[120] Craver 2015: 3.
[121] In favour of this proposal, see Wimsatt 1994, Eronen 2015, Eronen and Brooks 2018.

Craver, however, points out that these axioms generate several problems. Firstly, as for (i), it is neither really relevant nor useful to state that, for instance, a hippocampus is part of itself (for *reflexivity*). Secondly, as for (ii), *transitivity* does not work for functional parts, as noted by Humphreys too: "I am a functional part of the economy of Charlottesville, and my liver is a functional part of me, but my liver is not (yet) a functional part of the Charlottesville economy".[122] As for (iii), *extensionality* is a too weak criterion for object identity. In addition to its set of parts, an object is characterised by an internal organisation. Isomers in chemistry, for instance, are composed by the same parts, but their structure is different, and this difference produces different properties at the level of the whole. Eventually, regarding (iv), *summation* allows "to form arbitrarily many gerrymandered wholes out of disparate and unconnected parts with no spatial, temporal, causal, or functional unity"[123] and this is a practice of no utility in many disciplines such as in biology, where a whole is decomposed in functional parts that play specific causal and explanatory roles for the behaviour of the whole and that are "more or less isolable (nearly decomposable) sub-systems"[124] with a certain degree of autonomy.

In Craver opinion, therefore, classical mereology is a theoretical, formal theory which does not take into account those variables that are relevant to scientific practice. Here, space, time, and functions are taken into great consideration, so Craver recommends to integrate classical mereology with a further condition he calls *relevance condition*: "all the lower-level properties, activities, and organizational features of the parts are relevant to – contribute to – the property or activity of the whole".[125] The distinction between the part-whole relationship individuated by classical mereology and that identified by the integrated mereology formulated by Craver reflects the distinction between aggregation and organisation. Following Wimsatt, Craver states that aggregate properties are those for which the spatial, temporal and functional organisation of the parts is irrelevant. For instance, the mass of an object is just the sum of the masses of its parts; no matter how the parts are composed, all that counts are their individual masses. On the contrary, organised systems exhibit organised properties, namely properties for which the organisation of the components is relevant and influential to their nature. For instance, an organ such as a brain has complex properties due to its complex internal organisation, and a change in this organisation can

[122] Humphreys 2016: 17. About this see also Craver 2015: 14.
[123] Craver 2015: 15.
[124] Ibidem.
[125] Ibidem.

compromise the properties of the whole. On the one hand, aggregates exhibit invariance under the rearrangement of their parts, as well as under their substitution and re-aggregation – and the reason for this is that in aggregates parts do not interact among each other and the only properties that are relevant to the wholes are the individual properties of the components.[126] On the other hand, organised systems are integrated wholes in which the parts interact among each other in a complex network of relationships producing a sophisticated higher-level behaviour that is not just the sum of the individual behaviours of the parts. Craver uses the term "mechanism" to indicate "non-aggregative compositional systems in which the parts interact and collectively realize the behaviour or property of the whole",[127] and states that the division of natural phenomena into levels is possible just in the frame of particular mechanisms and not in more general global terms. That means, in other words, that levels of mechanisms are *local*. It is possible to say that the cells composing a hippocampus are at a lower level than the hippocampus, but saying that the cells composing a liver are at a lower level than a hippocampus is deceiving, because the cells of the liver have no part-whole relationships with the hippocampus, so the ordering is just conceptual and derives from a "generalisation over a relationship between tokens".[128] In other words, *if* in the levels metaphor the ordering relation is part-whole composition, then it is not possible to order into levels entities that are not related by part-whole relationships. Doing that just represents an abstraction, and the conceptual generalization that in nature there are monolithic, global layers of reality cutting horizontally nature and representing categories to which entities naturally and intrinsically belong is an abstraction as well. Levels are always defined in relation to a particular mechanism, so from an ontological point of view there are no absolute, comprehensive levels in which placing the entities of the world, but just particular levels correlated with particular mechanisms.

However, the absence of monolithic ontological levels does not imply that the decomposition of the systems we find in reality has to be inaccurate. For each organised system, structure, or mechanism, decomposition into levels is still possible, as well as it is possible to define the dependence relationships holding among the parts and the novelty the

[126] In Wimsatt, these points are formulated as follows: "**IS** (**I**nter**S**ubstitution) Invariance of the system property under operations rearranging the parts in the system or interchanging any number of parts with a corresponding numbers of parts from a relevant equivalence class of parts. (This would seem to be satisfied if the composition function is commutative). […] 3. **RA** (Decomposition and **R**e**A**ggregation) Invariance of the system property under operations involving decomposition and reaggregation of parts. (This suggests an associative composition function). 4. **CI** (Linearity) There are no **C**ooperative or **I**nhibitory interactions among the parts of the system for this property".
[127] Craver 2015: 16.
[128] *Ivi*: 18.

whole carry with itself. In this framework, however, the idea of a primitive fundamental level of reality uncorrelated to a specific mechanism or structure is excluded as well as any other absolute ontological level. Levels are local and piecemeal in being nested structures defined by different functions and mechanisms, and are correlated to the appearance of systemic constraints (we develop this point in §4.5). Their contextual nature, therefore, renders reductionism vain, because each decomposition just makes sense *per se*, without creating that sort of chain of being in which each level can be reduced to the level below.

3.3.4 *Levels and fundamental novelty*

In the previous paragraphs, we firstly saw that ontological bruteness and ontological fundamentality are related to ungroundedness and ontological independence. An entity is fundamental when it is not grounded on anything else and it does not depend upon anything else for its existence and essence. For these reasons, from an epistemological point of view, fundamental entities can be neither predicted nor explained in virtue of something else, being the truthmakers of themselves. Secondly, we noticed how emergence represents an interesting metaphysical case supposed to involve both dependence and fundamentality. Barnes suggests that emergents, because fundamental, should be considered at the same level of mereological atoms. Her view suggests, therefore, that there's one well defined fundamental level to which emergents and other fundamental entities belong. It is true that Barnes' is a conceptual, meta-ontological account taking fundamentality and derivability as the only relevant criteria for differentiation, but this account is disconnected by real problems of classification of entities, such as the fact that natural phenomena, as we saw, hardly offer the opportunity to create neat and precise distinctions among them and their structures. Emergents, moreover, appear in various domains of reality, especially in the most complex ones, therefore it seems at least controversial to place them in just one undifferentiated level, together with mereological atoms, too. Furthermore, emergence is characterised by fundamentality because of fundamental *novelty* and the relationship between levels and novelty is significant. It might be said that levels are defined by the appearance of genuine novel properties, so novelty might be viewed as their ontological mark: if fundamental novelty appears, a distinct level of organisation or being can be individuated. At the same time, the problem can be put in opposite terms, stating that the appearance or development of a new, distinct level of organisation produces genuinely novel features, being the latter a consequence of the nature of the former. In the first case, there are different levels of organisation because fundamental novelty appears at different scales and in correspondence of different relational structures or modes of organisation. In the second case, there is fundamental novelty in different domains and at different scales because of the different modes of organisation of matter. Now, these two formulations should not be necessarily viewed as mutually exclusive. New modes of organisation can realise new properties, which, in their own turn, can allow for new relational structures. What is clear, however, is that novelty is strongly correlated with the multiplicity of relational and organisational structures that are present in the natural world, rather than being localised at just one level of reality or correlated with just one kind of entity.

The idea of novelty characterising emergence is therefore an idea of scattered fundamental novelty appearing at different "levels" of reality – provided a piecemeal and local view of levels – and characterised by dependence upon an emergence base. Characterising emergence as fundamental can be misleading because the existence of emergents is grounded in and dependent upon relational structures or configurations of matter, so emergents are not really ungrounded, and not independent for their existence either. However, the properties emergents exhibit arise – and persist to exist – in coincidence with certain configurations of matter that can change or fluctuate in their composing parts. Emergent properties are *robust* upon these changes. They show autonomy, their features cannot be really traced back to their specific components – which do not show them – so, eventually, they exhibit original qualitative traits and new causal or determinative influence.

Now, fundamentality is often defined in reference to the origin or ontological basis of an entity: *x* is fundamental if its existence and essence is not produced, caused or derived by anything else, and for this reason God would have created it *ad hoc*. However, the kind of fundamentality here at issue seems rather connected with the presence of additional, original features in entities that *are* produced, caused or derived by something else. This derivation, though, does not exclude the appearance of features that are *different* from those instantiated by the components, and this difference is qualitative, not just quantitative, and reflects an ontological differentiation that can be also described through the idea of discontinuity in nature. These problems – qualitative novelty, differentiation, and discontinuity – will be approached in the next paragraph.

3.4 *Qualitative novelty*

Historically, the connection between emergence, discontinuity, and qualitative novelty took shape in the 19th century, in the framework of the philosophy of evolution.

As it is known, in formulating his theory of natural selection, Charles Darwin took distance from creationism and assumed a gradualist view of evolution, namely a hypothesis for which species (and their structures) evolve "slowly and successively"[129] over time through gradual and cumulative quantitative changes. The *quantitative* character of evolutionary change implied what Leibniz had defined the "Law of Continuity", which in Darwin is a fundamental principle often repeated in his works as "*natura non facit saltum"*:

> Why should not Nature have taken a leap from structure to structure? On the theory of natural selection, we can clearly understand why she should not; for natural selection can act only by taking advantage of slight successive variations; she can never take a leap, but must advance by the shortest and slowest steps.[130]

The canon *natura non facit saltum* is considered by Darwin "strictly true"[131] by his theory, and this implies that between all different forms of life there should be "numberless transitional links"[132] able to connect them in continuous chains of beings. The gaps found among species, moreover, should not testify in favour of their separate creations or sudden appearances, but are rather explained by the limits of our knowledge due to the incomplete exploration of the globe, the resulting imperfection of the geological record, and the fortuitous fact that just certain species have been preserved in a fossil state.[133] Therefore, even if it might be impossible to accurately outline the tree of life, this would be a contingent fact that should not implicate the ontological inference that no intermediate forms existed in correspondence to the gaps of our schemas. As stated in *On the Origins of Species*, "Nature is prodigal in variety, but niggard in innovation",[134] which means that in Darwin's view quantitative change regularly produces varieties, but no qualitative change producing innovation should be taken into serious consideration to explain the differentiation of species and the dynamics of evolution.

---

[129] Darwin 1859: 252.
[130] *Ivi*: 145.
[131] *Ivi*: 154.
[132] *Ivi*: 252.
[133] See chapter IX of *On the Origin of Species*, "On the Imperfection of the Geological Record".
[134] *Ivi*: 145.

Now, continuity was not just a weapon against creationism, but was supported by several facts highlighted by comparative anatomy too. The similarity and comparability noticed between structures belonging to taxonomically close species, such as the anatomy of orangutans, baboons and other primates, but also those between different groups, such as the wings of bats and those of birds, strongly suggested a sort of unitary plan in nature. New genetic facts discovered in subsequent years, moreover, would have further confirmed this similarity: the way in which genetic information is carried and transmitted by the DNA, for instance, is common to apparently very different forms of life, such as bacteria, animals and plants[135]. Darwin's gradualist account of continuity in evolution, however, incited some critical responses even among his contemporaries. Thomas Huxley, who was one of the strongest advocates of Darwin, stated as follows:

> Indeed we have always thought that Mr. Darwin has unnecessarily hampered himself by adhering so strictly to his favourite "Natura non facit saltum". We greatly suspect that she does make considerable jumps in the way of variation now and then, and that these saltations give rise to some of the gaps which appear to exist in the series of known forms.[136]

Leaps in continuity, in Huxley's opinion, were a real possibility, and ignoring them represented one of the problems of Darwin's theory. Huxley rightfully noticed that while continuity is perfectly acceptable to describe the development of structures, not the same can be said for the higher functions connected to them. In other words, a slight difference at the structural level (we would say, now, at the genotypic level), can produce a massive difference at the higher functional level (such as the phenotypic or behavioural one), so while continuity can explain the gradual differentiation of the former, it cannot really account for the surprising differences shown by the latter.[137] Those differences reasonably seem to represent discontinuous, qualitative changes taking place in the evolutionary process – at least at the function level, as we said.

Other authors did not accept Darwin's gradualism as well. Alfred Wallace, for instance, highlighted the evident differences between the inorganic and the organic, the vegetal and

---

[135] About these last points, see Ferrari 2015: 9 et seq.
[136] Huxley, 1864: 34.
[137] On this, see Blitz 1992: 36.

the animal, and the unconscious and the conscious. As for the last point, about consciousness, he dealt with a dilemma:

> There is no escape from this dilemma – either all matter is conscious, or consciousness is something distinct from matter, and in the latter case, its presence in material things is proof of the existence of conscious beings outside of, and independent of what we term matter.[138]

Rejecting panpsychism, Wallace eventually adopted a form of dualism admitting supernatural forces able to explain what he considered clear cases of qualitative novelty impossible to be explained through a mere accumulation of quantitative modifications. The supernatural character of these forces, however, were incompatible with Darwin's theory, so at the end of the nineteenth century, one of the authors that McLaughlin would have defined British Emergentists, Conway Lloyd Morgan, started to develop a theory of biological evolution intended to accommodate Darwinian gradualism and continuity, but also the presence of what seemed clear instances of qualitative novelty. The theory of emergent evolution, therefore, "seeks to interpret, on the one hand, the persistence and continuity of natural events, and, on the other hand, progressive advance with novelty".[139]

3.4.1 *Lloyd Morgan and the emergent evolution*

Lloyd Morgan was influenced by many authors who highlighted that even if the process of evolution is continuous and unitary, a good theory of evolution must take into account the presence of different domains of being. One of them was Wallace, who, as already said, traced distinctions between the inorganic and the organic, the vegetal and the animal, and the unconscious and the conscious. Another influential scholar who did something similar is Herbert Spencer, who stated that "there are three major spheres or stages within the evolutionary process: the pre-organic (the physico-chemical), the organic, and the super-organic (the psychological and social)".[140] Samuel Alexander too strongly influenced Lloyd Morgan, who outlined his synoptic schema for emergent evolution posing spacetime at the base, and Deity at the top, as done by Alexander.[141] What all these authors highlighted,

[138] Wallace 1870: 365
[139] Morgan 1923: 67.
[140] Blitz 1990: 156.
[141] For a detailed study about the influences underwent by Conway Lloyd Morgan, see Blitz 1990.

eventually, was an undeniable variety and ontological difference detectable in the nature of the diverse entities present in the world. The framework adopted by Morgan, therefore, accepted the existence of different *kinds* of entities belonging to different domains of reality and characterised by the instantiations of different kinds of "relatedness". As for the relations among all these different entities, Morgan recognised the distinction between resultants and emergents delineated by George Henry Lewes, for whom it was necessary to distinguish those properties "(*a*) which are additive and subtractive only, and predictable, from those (*b*) which are new and unpredictable".[142] The instances of this genuine evolutionary novelty in the course of history are countless: "Salient examples are afforded in the advent of life, in the advent of mind, and in the advent of reflective thought" ,[143] but Lloyd Morgan adds that also "in the physical world emergence is no less exemplified in the advent of each new kind of atom, and of each new kind of molecule".[144] Therefore, in Morgan's view, emergence represents the appearance of novel and unpredictable kinds of phenomena that show unprecedented relational structures and properties, even if completely composed by physical matter.[145] In fact, Lloyd Morgan highlighted that emergents are always co-occurrent (he said "co-existent") with resultants. Therefore, if Matter, Life and Mind are assumed to be three genuinely different domains of being, Mind will imply the existence of Life and Life will imply the existence of Matter, for there is no mental activity without physiological activity, and there is no physiological activity without physical-chemical events.[146] Morgan's formulation, therefore, is not supposed to introduce basic discontinuities in the development of nature, like a supernatural or a metaphysical dualist view would do. However, the passage from one domain to the other, and the emergence of new relationships and entities, is described as a qualitative "change of direction" of the process:

> Resultants give quantitative continuity which underlies new constitutive steps in emergence. And the emergent step, though it may seem more or less saltatory, is best regarded as a qualitative change of direction, or critical turning-point, in the course of events. In that sense there is not the discontinuous break of a gap or hiatus. It may be

[142] Morgan 1923: 3.
[143] Morgan 1923: 1.
[144] Ibidem.
[145] Carl Gillett recognises this kind of novelty as well: "(Qualitative Criterion) A property instance X, instantiated in an individual s*, is Qualitatively emergent if (i) s* is constituted of lower-level individuals s1–sn which instantiate properties P1–Pn, and (ii) X is a property not had by any of s1–sn, i.e. X is not identical to P1 or P2 or P3 or ... etc" (2016: 176).
[146] Ivi: 6 et seq.

said, then, that through resultants there is continuity in progress; through emergence there is progress in continuity.[147]

The novelty here at issue is associated with two elements. Firstly, entities belonging to different domains are "heterogeneous in the very nature of their being",[148] meaning that there are *intrinsic* differences between them (they have different qualities and different individual properties). Therefore, evolution produces "the outspring of something that has hitherto not been in being".[149] Secondly, lower and higher entities mutually influence each other, as lower-level entities group and form complex, higher-level compounds that acquire integrated unity thanks to novel emergent relationships. In these second cases, entities acquire new *extrinsic* relational properties connecting them to each other and to the world, and this new relational structures are *effective*, meaning that when they are present "some change in the existing go of events occurs, which would not occur if it [they] were absent".[150]

While the second cases seem to imply some sort of new and additional *causal* determination, the first cases, those of heterogeneity in nature, can be genuinely considered *qualitatively novel* as corresponding to both the appearance of entities that are novel because unprecedented in their own, and to salient qualities that were never instantiated before. Samuel Alexander, mentioning Morgan, for instance, said that the notion of emergence "serves to mark the novelty which mind possesses, while still remains equivalent to a certain neural constellation".[151]

3.4.2 *Qualitative novelty as innovation and differentiation*

Qualitative novelty, therefore, is a kind of novelty that is shown by certain phenomena even if those phenomena are composed by lower-level components and can be identified with them, as happens in Alexander. The qualities of the mind – *qualia* and semantics, for instance – do exist in their own forms even if mental states are realized by neural states and explicable by them. They do have peculiar characters depending for their instantiation upon their emergence base, but they are also heterogeneous from them in their essence. Here can be useful to recall Huxley distinction between structure and function. Even if structures

[147] *Ivi*: 5.
[148] *Ivi*: 14.
[149] *Ivi*: 112.
[150] *Ivi*: 20.
[151] Alexander 1920: 14.

evolve in a continuous way through slight, quantitative, physical modifications, the corresponding functions at the higher level can be surprisingly different in nature and can be able to change the course of an individual history. In biology this circumstance is particularly easy to notice, given the complex correlation between the genetic information stored in the DNA, which can be considered a sort of instruction for the development of the organism, and its phenotypical manifestations – its morphology, physiology, behaviour, way of interaction with the environment, and so on. Even if it might be possible to trace back a phenotypical trait to a particular genotype (ignoring the influence of the environment and all the other possibly relevant variables), the qualitative characters of the phenotypical trait would still be ontologically distinct in their nature and novel in respect to the set of genetic instructions coding for them. In these terms, evolution produces a huge amount of qualitative novelty – namely, innovation – and this is the reason why scholars such as Lloyd Morgan, Roy Wood Sellars,[152] but also contemporary philosophers such as Mario Bunge, considered evolution a process intrinsically involving the phenomenon of emergence.

The focus all these authors, together with many others, put on qualitative novelty highlights another interesting point, namely that the nature of reality has to do with differentiation and discontinuity, rather than with unification. Emergence is not a metaphysical dualist view. It implies substance monism, which means that everything in the world is constituted by just one kind of matter, which is physical. Each entity shares with all the other existing ones the property of being physically composed, so even emergents, as well as non-emergents, are made of physical stuff. However, commonality of composition can just partially explain the nature of most of entities, as their components can be organised in different ways and organisational differences are causally salient. Moreover, another feature that is essential to the identity of an entity is the context or environment in which it exists. Therefore, to understand the structure of reality and the properties of different natural phenomena, searching for their common compositional parts is not exhaustive. As stated by Sandra Mitchell: "Compare a conglomeration of molecules constituting a rock with an organization of molecules making up a baby monkey. What's the difference? It is not to be found by looking at what they share, namely physical composition, but only by looking at how they differ".[153] In simple words, differences in organisation and interaction with one

[152] See Sellars 1922.
[153] Mitchell 2012: 178.

another, as well as with the context/environment, are far more essential than their being composed by the same microphysical elements.

The notion of emergence implies the importance of all these additional features for defining and identifying complex entities. Emergent phenomena are composed by the same kind of components composing non-emergent phenomena, but their internal organisation allows for the appearance of distinctive qualitative features, as well as new modes of interaction and novel and sophisticated behaviours which the components do not show. For many British Emergentists, eventually, this kind of qualitative, higher-level novelty was the one emergence implied, for it was a kind of novelty which did not rule out the possibility to decompose without residue each entity into physical components, and this point was one of the aim of emergentism, namely to acknowledge for novel entitites without accepting dualist or supernatural accounts of reality. Qualitative novelty responded perfectly to this need and still does, if we think about how – in a very different field and about a century later – the same theoretical strategy is adopted by some philosophers of physics who are concerned with determining the origin of a complex phenomenon such as space-time. Also in this case, as we will see in the next paragraph, the phenomenon to be explained is defined at the same time as both reducible and emergent, where the keystone of this double characterization corresponds to qualitative novelty.

3.4.3 *Emergent spacetime*

Space and time have always been considered essential elements of any fundamental ontology, so much so that to exist physically has been and still is normally understood as existing in space and time, and to exist in general seems to require at least existence in time. Yet contemporary physics seems to suggest something different.

It is well known that general relativity is currently the best theory of gravity and space-time at our disposal. Nevertheless, this theory does not really seem fundamental because it fails «at extremely short length scales (corresponding to high energy scales), and in regions of extremely high curvature, where quantum effects cannot be neglected»[154]. Given these limitations, in recent years scientists have attempted to formulate a quantum theory of gravity that would provide a more complete description, and most research programmes investigating quantum gravity domains describe them as non-space-time domains, implying

[154] Crowther in Onnis 2019: 76.

that space-time is not a fundamental element of reality, but on the contrary is derived from more fundamental non-space-time structures.[155] In this framework, space-time has been interpreted as an emergent phenomenon, but, as we will see shortly, the definition of emergence exploited in this case is anomalous compared to the widespread standard vision investigated in philosophy, but also compared to what seems to work for other physical phenomena, such as those described by Humphreys.

First of all, the *relata* of the emergence relationship can be both the theories (quantum gravity on the one hand and the theory of relativity on the other) and the physical structures whose existence is postulated by these theories (note that most philosophers of physics share a naturalised metaphysics that commits them to the existence of what is postulated by the best available physical theories). In both cases, the relationship is asymmetrical: on the one hand, it is configured as a basically derivative relationship between a more fundamental theory and a less fundamental one, in the sense that the less fundamental theory (relativity) must be derivable from the more fundamental one (quantum gravity); on the other hand, it is a relationship of dependence that links the most fundamental non-spatio-temporal structures to the emergent spatio-temporal structure, which is realized (albeit in a multiple way) by its emergence base.

The first characteristic that is recognized to the theories of space-time and the phenomenon itself is therefore their *dependence* on a more fundamental theoretical or ontological basis: a criterion that is certainly widespread, but not always expressed in terms of derivability and reductiveness (we will return to this point shortly). The second distinctive feature of the emergent space-time is instead represented by a specific type of *novelty*, defined by Wüthrich in terms of *qualitative distinction*.[156] In other words, the emergent spatio-temporal universe and the non-spatio-temporal universe that represents its emergence base are significantly different because of the non-spatio-temporality of quantum gravity structures: this non-spatio-temporality has the consequence of making these fundamental structures devoid of dimension, duration, reciprocal distance and so on, characteristics that the emergent, novel spatio-temporal structures instantiate instead.[157] As Crowther puts it, «Novelty is taken as robust behaviour exhibited by the macro-system (appropriately described by the emergent theory) but not present in the micro-system […] The emergent

[155] See Huggett & Wüthrich, 2013, Wüthrich in Gibb, Hendry & Lancaster 2019.
[156] Wüthrich in Gibb, Hendry & Lancaster 2019: 318.
[157] See Wüthrich in Gibbs, Hendry and Lancaster 2019 and Crowther in Onnis 2019.

theory is novel compared to the theory it emerges from if it is formally distinct from the latter, describing different physics and different degrees of freedom»[158].

Secondly, the spatio-temporal physics that emerges from the lower level domain manifests robustness and autonomy because the relationship of dependence between the fundamental and the emergent structure is configured in terms of multiple realizability and universality. For this reason, the dependence in question is relatively weak and the emergent physics is therefore underdetermined by the fundamental one. We can therefore conclude that the concept of emergence in the domain of quantum gravity and relativity refers to a relationship of simultaneous dependence and independence between fundamental and emergent structures. This relationship, however, is seriously qualified by another property: that of exhibiting qualitative novelty. This novelty is indeed a precondition for other forms of novelty, but has its own individual relevance in the fact that it represents the appearance of structures that have never existed before or that bear new causal powers, but which are intrinsically different from everything else that exists.

3.4.4 *Qualitative or epiphenomenal novelty?*

I would like to add a final remark about the confusion that often arises between *qualitative* and *epiphenomenal* novelty. Thomas Huxley, who brilliantly noticed the distinction between structures and higher-level functions, had an epiphenomenal view of consciousness, which should be considered, in his opinion, just "[…] a collateral product of its [the brain] working, and to be completely without any power of modifying that working as the steam-whistle which accompanies the work of a locomotive engine is without influence upon its machinery".[159] So, in Huxley's view, even if higher-level functions should be recognized as distinct from lower-level structures, no causal influence on lower-level physical structures should be acknowledged. In fact, Huxley considered animals, as well as human beings, as machines, but, significantly, *sensitive* and *conscious* machines,[160] where matter exerts causal powers on the mind, but not the other way around: " […] all states of consciousness in us […] are immediately caused by molecular changes of the brain-substance. It seems to me

[158] Crowther 2016: 52.
[159] Huxley 1874: 240.
[160] *Ivi*: 238.

that in men […] there is no proof that any state of consciousness is the cause of change in the motion of the matter of the organism”.[161]

This last quotation expresses the central problem of the mind-body question, that is, how they interact. According to Huxley, there is a one-way relationship between them: the body has a causal power over consciousness or the mind, but there is no evidence to the contrary. However, according to a rather traditional view, if consciousness cannot directly cause the motion of matter it is because it is not itself matter. Causality has in fact long been defined and conceived as a relationship that links two events in a physical process, as happens in the most conventional example of causality, i.e. the one involving collisions between spheres (often represented in the form of billiard balls). The problem of defining the type of relationship connecting one event to another in causal terms has raised a wide debate in the literature and various proposals and interpretations have been made throughout history. All these models of causality, which interpret the causal relationship as nomological subsumption, statistical correlation, contiguous changes or energy flows (to name a few)[162] can be traced back to two main models: the one that sees the connection between cause and effect as a formal statistical relationship, where the occurrence of the cause makes the occurrence of the effect more likely, and the one that sees this connection as a real exchange of energy or a process involving mechanical or physical forces, so that when a causes b actual exchanges of matter or energy come into play.[163]

Now, the first interpretation of the causal connection does not tell us anything about the nature of the *relata*, this being a formal relationship. The second interpretation seems instead limited, since it admits causality only in the case where there are direct passages of matter and energy between entities that must be strictly physical[164]. The debate on the appropriateness of the latter theory is heated and does not seem able to provide necessary and sufficient conditions for causal relations.[165] As regards the debate on emergence, instead, this interpretation seems to make it impossible a priori to attribute causal effectiveness to complex entities, since the criterion for doing so depends on the ability to directly exert physical forces that are always exerted by the constituent physical entities. The suggestion I

[161] *Ivi*: 244.
[162] See Schaffer 2016 for an overview.
[163] On this, see Schaffer 2016 and the next chapter.
[164] See, for instance, Max Kistler: “Two events c and e are related as cause and effect if and only if there is at least one physical quantity P, subject to a conservation law, exemplified in c and e, of which a determinate quantity is transferred between c and e” (Kistler 2006: 9)
[165] Anscombe 1975.

will make in the next chapter is that this vision should be reviewed and expanded because it is arbitrary and unjustified. As we shall see shortly, the complex phenomena that manifest emergent properties seem to exert some kind of determination in binding the behaviour of the components of their emergence base, and this determination is real and must be taken into account even if it does not fall within the scheme of causality understood as a direct exertion of forces. It seems necessary, at this point, to analyse the distinction between direct causality – normally interpreted as a physical process – and these other forms of constraining determination that could characterize the phenomenon of emergence. The next chapter is dedicated to this.

CHAPTER IV
Causation and determination

4.1 *Introduction*

As we noticed in the first chapter, different taxonomies describe emergent phenomena in different terms, but the three representative authors we took into consideration, as well as many others they mentioned, acknowledge a fundamental distinction between a first kind of ontological emergence representing a genuine feature of reality, and a second kind of epistemological emergence that can be viewed as a consequence of our epistemic limitations or insufficient computational power in the analysis of reality. As for the first kind of emergence, which has been defined as *strong*, *ontological*, or *metaphysical*, there is one essential requirement almost universally recognised, namely the presence of novel, irreducible causal powers, allowing, sometimes, for downward causation.[1] Jessica Wilson is one of the most representative cases of this trend. As shown in §1.4, Wilson focuses on special science entities and states that despite being materially composed and partially dependent upon complex configurations of more fundamental entities, they exhibit a degree of ontological and causal autonomy, being "[…] distinct from and distinctively efficacious as compared to the micro-configurations upon which they depend".[2] In Wilson's view, strong metaphysical emergent entities exhibit novel and distinct causal powers not present at lower-levels of reality and this implies the rejection of the Physical Causal Closure premise, which states that "every lower-level physically acceptable effect has a purely lower-

[1] Among the authors who express this requirement, we can mention O'Connor 1994, Bedau 1997, Chalmers 2006, Gillett 2016, Humphreys 2016 and Wilson forthcoming.
[2] *Ivi*: 2.

level physically acceptable cause".[3] In her view, therefore, if strong emergence is accepted, then the domain of the physical is not nomologically complete and closed.

Now, in the framework delineated by Wilson, causation is based on *exerting powers*, and the possibility of higher-level causal efficacy is based on the possibility of bestowing new powers or new power profile. As it is well known, higher-level causal efficacy has been widely criticised. Recently, two significant kinds of criticism have been defined the two faces of the "collapse objection". This objection, however, seems effective in the framework of this particular account of causation based on bestowing powers. In order to neutralize the objections, therefore, it may be effective to revise the notion of causation. Rather than being just definable as the capacity to exert powers, causation may be reread in wider terms so as to include other forms of "non-causal" – as they are often defined – determinative and constraint relations, for which the classical objections do not hold. In the next paragraphs, after presenting the collapse objection and some relevant answers, we will explore the field of those "non-causal" relationships that are able to "make differences" in the world, delineating an extended view of causation partially inspired by the so-called *difference-making* account but substantiated by metaphysical concerns.

4.2 *The collapse objection*

The possibility for higher-level causal efficacy has been widely criticised in – at least – two different ways that have been recently defined by Jessica Wilson and Umut Baysan as two versions of the "collapse objection".[4] According to Wilson and Baysan, the first, classical version traces back to Jaegwon Kim's *causal inheritance principle*. Kim noticed that any emergent phenomenon must have an emergent physical base, so there is no reason to exclude that the physical base is the bearer of all the causal powers the system as a whole exhibits: "if an emergent, *M,* emerges from basal condition *P*, why cannot *P* displace *M* as a cause of any putative effect of *M*?".[5] The causal inheritance principle, in other words, states that the causal powers of the higher-level features – mental properties, in Kim's formulation – are inherited from the lower-level physical features:

[3] Wilson forthcoming: 56.
[4] See Taylor 2015, Baysan and Wilson 2017, Wilson forthcoming.
[5] Kim 2006: 558.

[The Causal Inheritance Principle]: If mental property *M* is realized in a system at time *t* in virtue of physical realization base *P*, the causal powers of *this instance of M* are identical with the causal powers of *P*.[6]

If the principle is true, then "any supposedly novel powers of S [*a strongly emergent feature*] will 'collapse' into those of P [*the corresponding physical base*]".[7] Therefore, what a system can be and what a system can do in certain conditions always depend upon its lower-level physical features *alone*, and no space is left for causally efficacious emergent properties.

The second version of the collapse objection relies on the notion of *dispositional property*. A dispositional property is a property identified by the causal and functional powers it confers to the entity in which it is instantiated. A classical – although controversial, as we will see – example of a dispositional property is fragility. An object is fragile when, given certain conditions or "stimuli" (and given certain laws of nature), it breaks. So, in other terms, dispositional properties imply counterfactual conditionals: for a glass, to be fragile means that *if* the glass is hit (stimulus), *then* it will break (manifestation). This way to describe dispositional properties is called *counterfactual analysis* (CA) and can be formulated – and formalised – as follows:

x is disposed to manifest M in response to stimulus S *iff* were *x* to undergo S *x* would yield manifestation M […]

(CA) $D_{(S,M)}x \leftrightarrow Sx \rightarrow Mx$ [8]

This model is very intuitive, but it is not faultless because it is refuted by numerous cases in which an entity with an M disposition does *not* manifest M despite the occurrence of conditions S. This is due to various circumstances involving so-called "finks", "mimics" and "masks" (or antidotes). In the first case, a finkish disposition disappears despite the occurrence of stimulus S due to the presence of a so-called fink. An effective example is the one provided by C.B. Martin in a 1994 article entitled *Dispositions and Conditionals*[9]. Imagine a metal wire that has the disposition to conduct electricity. The conditional analysis of this example would be: "if the wire were to be touched by a conductor, the electric current

[6] Kim 1993: 326.
[7] Wilson forthcoming: 168.
[8] Bird 2007: 24.
[9] Martin 1994.

would flow from the wire to the conductor”.[10] However, if the wire was connected to a mechanism that made it inert whenever it received electricity (this mechanism would be the fink), then the wire would *not* conduct electricity despite its disposition to do so. This counterexample shows that the conditional analysis of the dispositional properties does not always hit the mark, since in this case the conditional is satisfied (S occurs), but disposition M does not occur.

In case of *mimics*, a certain circumstance emulates the effects of a disposition and produces the same counterfactual even if no dispositional property is really present. An example can be a robust (not fragile) pot connected to a sensitive detonator. In this case, when the pot is hit it will explode and fall apart, despite not being fragile at all. Here, the conditional “S*x* → M*x*” would be false because we would have M without the occurrence of S. In the third case, that involving maskers or antidotes, the disposition is maintained but an interfering event changes the causal chain connecting the instantiation of the stimulus S to the manifestation M. For instance, in a case of poisoning (M) after the ingestion of a toxic substance (S), an antidote taken before S or immediately after it would block the manifestation of M. Here, the disposition is not neutralised as in the first case, but an external causal factor creates an interference which renders false the conditional “S*x* → M*x*”. The topic is clearly complex, so we refer to Stephen Mumford’s *Dispositions* and to Alexander Bird’s *Nature’s Metaphysics* for an in-depth analysis.[11] However, what is clear is that dispositional properties are properties that imply certain potential effects that will be produced in the future once given the right conditions: “To ascribe a disposition is to suggest possibilities of behaviour. It is to say that something could or would happen if the conditions were right”.[12]

Non-dispositional properties, on the contrary, are called “categorical” and do not depend upon anything external to the entity itself. For instance, an object[13] is (or is not) triangular regardless other objects, events or contextual conditions. While *being fragile* is a dispositional property because it depends upon certain conditions for the production of certain potential effects, therefore, *being triangular* is a categorical property in being just

---

[10] Marmodoro & Mayr 2019: 73.
[11] Mumford 2003, Bird 2010.
[12] Mumford 2003: 5.
[13] See Gozzano

defined by itself and in having "no necessary connections with other entities"[14] to be what it is, and no "intrinsic causal significance".[15]

Now, dispositional properties can represent a threat for emergence because it may be affirmed that even if some properties that emergent entities exhibit are not instantiated at the lower-level of their components, the components could still possess these features as hidden or latent dispositions appearing at the higher level of organisation. Eleonor Taylor, in analysing C.D. Broad's example of the chemical compound sodium chloride, for instance, states that the features of this chemical whole – let's say *x*, *y*, and *z* –, which Broad considered impossible to be deduced from the complete knowledge of the features of the components, i.e. sodium and chlorine, can indeed be deduced if the dispositional property of '*generating a compound characterised by* x, y, *and* z' is included as a dispositional property among the individual properties of sodium and chlorine. In other words:

> The emergent features of the whole R (A, B, C) can obviously be deduced from complete knowledge of the features of the parts A, B, and C and the knowledge that they are arranged as a whole R (A, B, C), *so long as the features of the parts include these dispositional properties* [emphasis mine].[16]

Once dispositional properties are included among the properties of the lower-level components, therefore, emergence collapses and disappears, and this is because emergence implies the presence of at least two different sets of properties: lower-level properties and higher-level properties. If there is no distinction among these two groups of properties and all the properties are lower-level ones, then no property seems left to be emergent.[17]

In the next paragraph possible answers to this twofold objection are illustrated.

---

[14] Bird 2010: 67.
[15] Dumsday 2019: 1.
[16] Taylor 2015: 736.
[17] Another formulation of this objection can be traced back to Shoemaker, when he states that lower-level properties can have both "micro-manifest" and "micro-latent" powers. While the former are manifest, the latter become manifest just when "combined in an emergence engendering way" (Shoemaker 2002: 55). However, these powers becoming manifest only in certain conditions, despite being emergent are not strongly or metaphysically emergent, in Shoemaker's view, because they have no novel fundamental powers not had by their physical constituents.

4.3 *Answers to the collapse objection*

Now, as for the first version of the objection, causal inheritance depends upon the particular relationship holding between the entities belonging to the lower-level and those belonging to the higher. This relationship, canonically, is the realization and in cases where it is involved the principle applies to the functional properties, which are precisely realized. According to Wilson and Baysan, however, the principle is sometimes applied also in cases where the entities belonging to the two levels are linked by a nomological need. Imagine a phenomenon S emerging from the physical base P to which the effect of producing P* is attributed. If causality is understood in terms of nomological need and therefore it is enough that P is nomologically sufficient to have M to affirm that P is the cause, then, by virtue of the transitivity of necessity, if P is sufficient to have M and M is sufficient to have P*, P is sufficient to have P*. This means that if the higher-level property M is nomologically needed by the physical base P, then the causal powers of M will be the same as P or a subset of it.

However, this version of the collapse objection does not appear to be solid for a number of reasons. The first depends on the fact that the interpretation of causality as a nomological necessity tends to be more complex than described. Usually a fact (such as the instantiation of P*) is caused by a set of causes, rarely by a single cause, be it the instantiation of M or P. This linear causal pattern, whereby cause *a* is followed by cause *b*, regardless of context and consideration of what came before and what will come after, is most often inadequate because it is too simplistic. As we have already said about the realization model formulated by Shoemaker, let's imagine this series of facts, which seem to be causally linked. The stimulation of the C fibres (P) produces a sensation of pain (M) and this pain generates a feeling of sadness that will certainly have physical realizers (P*). If the instantiation of P is nomologically sufficient for the instantiation of M and the instantiation of M is sufficient for the instantiation of P*, then, for the transitivity of necessity, the instantiation of P will be nomologically sufficient for that P*. Yet it is clear that neither P nor M are sufficient, on their own, for P* to be given: in reality, other contextual conditions are always necessary which form a set of causes whose composition may be more or less heterogeneous. Let us suppose that I bet a considerable amount of money on the fact that I would have a headache on such a day at such an hour as a result of some stunt I pulled with friends. If in those conditions P and therefore M had actually been given, P* would not have followed, but the exact opposite would occur because I would have been in pain, but happy.

Defining small fragments of causal chains in terms of nomological necessity can therefore be inconvenient[18] and it should be noted that this type of definition does not provide metaphysical explanations of the causal relationship, but merely ascertains systematic regularity by defining them modally (i.e. defining them as necessary).[19] An important premise of the first version of the collapse objection, therefore, may be subject to criticism, since it is not obvious that causality should be interpreted as a mere nomological necessity,[20] and even if it were reasonable to do so, a definition that takes into account sets of causes, rather than individual causes, would seem more appropriate.

In any case, the most convincing objection to the first version of the collapse was made, in my opinion, by Baysan and Wilson.[21] They point out that the powers exercised by the higher level entities are exercised at the higher level and not at the microscopic level of the components. Therefore, a response to the collapse objection could focus on the distinction between *direct* and *indirect* powers. In the case of an ontologically emergent phenomenon, the emerging powers are not possessed or manifested by the lower level entities in the same direct and immediate way in which they are possessed or manifested by the emergent entity[22]: the exact *locus* of the emergent power remains the higher level, even if the physical base nomologically necessitates it. The physical basis, according to Wilson and Baysan, should therefore not be understood as what actively exercises power, but as its precondition or its precursor. Imagine this scenario suggested by the authors, which does not exploit relationships of realization, but facts linked by nomological necessity: a person P lights the fuse of a bomb and causes an explosion. The fact a (the lighting of the fuse by P) nomologically requires the fact b (the explosion), because if the fuse had not been lit by P, the explosion would not have occurred (*ceteris paribus* - i.e. all other conditions being equal). Now, this chain of causal events does not allow us to attribute to P the *direct* and *immediate* causal power to produce an explosion. In other words, P could not detonate anything without a bomb. The power to produce an explosion therefore belongs to the bomb and not to P, even if P is its precondition and forerunner because without anyone to light the fuse, the bomb remains inert and does not manifest its explosive power. Likewise, the fact that the physical

---

[18] On this see Kment 2010.
[19] On the fact that modal connections are inadequate for metaphysical analysis, see Fine 1994.
[20] At the end of the third chapter we mentioned that causality models can be divided into two groups, those who see the connection between cause and effect as a formal statistical relationship and those who see it as a physical process involving the exchange of energy. The type of interpretation of causality that resorts to the notion of nomological necessity belongs to the former.
[21] Wilson and Baysan 2017, Wilson forthcoming: 173 ff.
[22] Wilson a breve: 173.

base P nomologically needs the P* effect of M does not imply in any way that the P* effect is produced directly by the powers of P, since, as already pointed out, no metaphysical conclusions can be drawn from purely modal analyses.

The second version of the collapse objection, conversely, addresses the metaphysical problem of how and why emergent properties would not exercise genuinely new and original powers. These powers would be exerted by the dispositional properties of the microconstituents, which would remain latent until the right stimulus conditions occur. However, criticism can also be raised with regard to this objection.

First, as suggested by O'Connor, the objection might be trivial. O'Connor states that despite the impossibility to clearly dismiss the dispositionalist move, it seems implausible for the following reasons:

> Why does such a micro-property make its presence known only in highly complex systems of a certain sort? How is it that such a fundamental property can be so causally isolated from other micro-properties so as to be discernible only in circumstances that are otherwise noteworthy only for the complex macro-properties which are instantiated? […] the only motivation one could have for postulating a (rather elusive) micro-property is a very strong methodological principle to the effect that one is to avoid emergentist hypotheses at all costs, which by my lights is not a reasonable one.[23]

In other words, in O'Connor's opinion the collapse objection based on dispositionalism may reflect an apriori anti-emergentist methodological attitude. His conclusion is based on the idea that a fundamental micro-property could hardly be manifest just in complex, high-level organisations, therefore the decision to "anchor" the property to the micro-level is not a motivated move, but just the mark of a prescriptive metaphysical preference that deliberately avoids emergentism. However, dispositional properties, as pointed out by Baysan and Wilson, always become manifest just in certain particular complex structures, for to exhibit certain properties, the occurring of certain conditions appears to be necessary. O'Connor's reluctance, therefore, may not be justified by this reason. Nonetheless, his remark might still be correct because rejecting emergent properties in favour of latent dispositional properties can be arbitrary for other reasons. As explained by Neil Williams, the ascription of a disposition does not constrain to a particular ontological view:

[23] O'Connor 1994: 98-99.

"Disposition ascriptions are ontologically neutral"[24] because "asserting of a glass that it is fragile commits us to the existence of whatever truthmakers are needed to make that assertion true, but it does not yet commit us to any specific ontology"[25]. This is the reason why Williams builds his metaphysics focusing on what grounds dispositions, rather than on dispositions *per se*. Those grounding elements are powers, and the author states that focusing on powers and not on dispositions is a wise choice based on the idea that the term "disposition" can generate relevant confusion. In Williams' opinion, an appropriate powers metaphysics implies that entities manifest dispositions *because of* powers, which once arranged in particular "constellations", manifest certain unprecedented effects and abilities. Dispositions, conversely, are often intended as tendencies, and the attribution of a tendency is in Williams' opinion metaphysically neutral, because the tendency for a dispositional ability "requires that the object's ability has been exercised often enough and reliably enough that we are able to track it. It is, therefore, a historical notion, perhaps even an epistemological one".[26] On the contrary, grounding dispositions into powers allows for a precise metaphysical view in which powers are fundamental and able to produce dispositions at the level of the objects that had those powers (it is worth noticing that in Williams view, objects have properties and properties are powers[27]). In this framework, for instance, the disposition *being fragile* belongs to an entity such as a glass because of the molecular structure of the object – i.e. its powers. Given the powers had by the glass, the glass has the disposition to break in certain conditions. Therefore, at the fundamental level there are certain constellations of powers, and at a higher level there are the dispositions produced by those constellations and possessed by the objects. Alexander Bird's view of dispositions, which he defines *potencies*, highlights the same point. Dispositions are not fundamental; rather, they are realized by certain causal basis and mechanisms placed at a deeper level; consequently, the identification of a disposition requires "knowledge of a deeper level of science",[28] and this knowledge is essential to understand what it is to have that disposition. One of Bird's examples is the disposition of *being poisonous*. This disposition can be realized in many different ways:

[24] Williams 2019: 55.
[25] Ibidem.
[26] Williams 2019: 49.
[27] See Williams 2019: 46
[28] Bird 2010: 40.

> So a lot of different substances might be poisonous and poisonous for different reasons. Some might be neurotoxins, some might interfere with a crucial metabolic pathway or cause a malfunction in any of the body's vital organs. And a poison might do any one of these things in a wide variety of different ways.[29]

For Bird too, therefore, disposition ascriptions alone cannot shed light on the particular ontological realisers or producers responsible for their manifestations, and this vagueness seems hardly able to rule out the presence of higher-level properties or powers. Moreover, dispositions appear to be non-fundamental properties or powers because of their nature of realized or produced phenomena. The distinction between two different levels of properties or powers, therefore, might still be valid, differently from what stated by Taylor, for whom emergents should be dismissed given the collapse of their alleged level into the micro-physical one.

A further remark might be, moreover, the apparent artificial nature of those properties Taylor recognises as "latent dispositions". Taylor takes into account three candidates of emergent properties and illustrates her collapse strategy for each of them. The first is the already mentioned case taken from Broad about the properties of the compound sodium chloride, which are converted by Taylor into the latent dispositional property had by each individual part: '*generating a compound characterised by x, y, and z*'. The second case is taken from Bedau, who attributes weakly emergent properties to some *Game of Life* configurations, such as the R-Pentomino. The property highlighted by Bedau corresponds, in Taylor's view, to the latent dispositional property '*forming a configuration whose development halts at 1103 generations when combined with other cells into an R-Pentomino'*. The third case regards Chalmers' theory of consciousness, which is emergent because of the failure of a logical necessitation by the physical laws. In Chalmers view, the relevant laws connecting physical facts to qualitative experience are a particular kind of psycho-physical laws, but in Taylor's opinion the property of '*being governed by the Psycho-Law*' can be easily attributed to the microphysical components as one of their latent dispositional properties.

Now, these three latent dispositions are complex, extrinsic properties whose manifestations appear at a very complex level, and they seem elaborated *ad hoc*, with the purpose of ruling out the emergent properties recognised by the authors Taylor contests. It

[29] Bird 2010: 39.

is not clear why these higher-level manifestations should be attributed to lower-level properties, however. This attribution is neither impossible, nor improbable, but nor straightforward as declared by Taylor. The manifestations could be related to lower-level properties and/or to higher-level relational properties too. As pointed out by Williams, without an appropriate metaphysical analysis able to ground manifestations and dispositions in some way, the attribution of a disposition remains at a superficial, observational level. Taylor does not satisfactorily specify *how* a compound like Broad's sodium chloride should be attributed of a disposition such as '*generating a compound characterised by* x, y, *and* z'. What she says is that if we "broaden the micro-level properties to include the extrinsic properties", we could include those latter properties into the set of the micro-level properties. It remains to understand, however, why we should do that. First, this would render every more or less arbitrarily individuated extrinsic property a micro-level intrinsic property, regardless its features and locus of manifestation. This would render the debate trivial, and in fact this objection traces back to the nineties, to van Cleve.[30] Secondly, dispositional properties are hardly considered extrinsic.[31] An extrinsic property is a property "that is non-intrinsic, that is, when its being had by an object is in virtue of objects or properties outside the object in question".[32] The Intrinsicality Thesis for dispositions, by contrast, is almost universally accepted: dispositional properties are intrinsic, their instantiations depend upon the object in which they are found, and they produce their manifestations in virtue of certain conditions that involve other properties or relations. Therefore, manifestations depend upon other entities, but the fact that an entity *has* the dispositional properties it has does not depend upon other entities but just upon its intrinsic properties (causal basis or powers). Dispositions are intrinsic and actual properties, and when they are latent, to be latent are just their manifestations. Dispositions are always instantiated and always actual, even if their effects are not manifest.

Another valuable objection to Tayler's collapse strategy is the one formulated by Baysan and Wilson.[33] They notice that higher-level properties exhibit specific powers, and those powers are exerted at the higher level of the compound, and not at the microscopic level of

[30] Van Cleve 1990: 222-3 "[…] is it not true of sodium that it comes with chlorine to form a whole having such-and-such properties, including its odor and anything else one might have claimed to be emergent? And from such properties of the parts, may not all properties of the whole be deduced? The answer, of course, is yes; but it is also clear that if properties of this sort are admitted in the "supervenience base," the doctrine of anti-emergence (or mereological reducibility, as it might be called) becomes completely trivial".
[31] See Mumford 1998, Heil 2003, Molnar 2003, Bird 2010, and Williams 2019.
[32] Williams 2019: 67.
[33] Wilson and Baysan 2017, Wilson forthcoming: 173 et seq.

the components. Therefore, an answer to the second version of the collapse objection can be focused on the distinction between direct and indirect powers. In their view, in the case of a Strong emergent *S*, *S*'s purportedly novel powers "are not had or manifested by lower-level features in the same direct or immediate way as they are had or manifested by *S*".[34] What Wilson is saying here, is that the precise *locus* of the emergent power remains *S*, even if the physical base instantiates a dispositional property and consequently necessitates *S* as its precondition or precursor. In this sense, having a dispositional power would correspond to being a precondition or a precursor of a power that will be directly exerted by another feature (in this case *S*) at another level of organisation. Why this view should rule out the possibility of emergent properties, however, is not clear, for emergent powers could be the manifestations of the potentialities of the components, when those potentialities are actualised by the instantiation of the right conditions or stimuli. These manifestations, as pointed out by Baysan and Wilson, have powers too, and this is not inconsistent with dispositionalism. If we accept pandispositionalism, then manifestations would be powers as well, so a particular scenario in which different properties exert different powers at different levels of powers composition is produced again, *contra* the suggestion that latent dispositions rule out higher-level properties. Dispositions are produced or realised by lower-level powers, but they manifest, given the right conditions, higher-level powers: "Dispositions manifest themselves in properties",[35] and properties are *clusters* or *network*s of powers.

Assuming the possibility of emergent properties in this framework is still feasible, and actually done by Mumford and Anjum, for instance, who recalling John Stuart Mill, considered emergent properties as the clusters of powers produced by a "surprising, nonlinear, non-additive"[36] composition of other powers. In these cases, powers are composed in a way that the resultant effects cannot be understood but *holistically*, for their joint causal manifestations "cannot be reduced to the isolated workings of the individual powers involved".[37] Mumford and Anjum conclude, therefore, that

[34] Wilson forthcoming: 173.
[35] Mumford & Anjum 2011: 5.
[36] *Ivi*: 87.
[37] Ibidem.

> This suggests some variety of emergentism because two powers could produce a novel phenomenon that can be both surprising (weak emergence) but also genuinely productive of something new in the world (strong emergence).[38]

To conclude, the second version of the collapse objection relies on the idea of latent dispositions, and suggests that emergent properties can be ruled out by the presence of latent dispositional properties had by the microphysical components. The latter, in Taylor's view, would be responsible for the powers allegedly exerted by the former, making emergents superfluous. Dispositionalism, in Taylor's view, is a huge threat for emergentism, and she suggests that a viable defence against this menace would be the rejection of the possibility of dispositional micro-level properties: "The emergentist could stipulate that the micro-level properties in cases of emergence are exclusively non-dispositional properties, and thereby hope to avoid the collapse problem".[39] However, the idea that dispositional properties are inherently able to rule out emergent properties is not conclusive, as we saw. First, dispositions attribution admits different metaphysical scenarios; second, dispositions grounded in clusters of powers can themselves produce novel properties and novel effects that can be viewed as emergent. The collapse objection, therefore, does not appear to be conclusive, and still leaves space for the possibility of strong emergent properties exerting novel determinative or causal powers.

4.4 *An extended view of causation*

As stated by David Yates, "in general, properties are causally relevant by dint of the causal powers they bestow".[40] The classical view of causation, namely, states that being causally relevant *corresponds to* genuinely and directly bestowing active causal powers, whereas inheriting causal powers from other entities – such as one's own realisers – does not mean being fundamentally causally relevant, but just being derivatively so (which means, in fact, not being *really* causally efficacious). Now, reality is composed by different kinds of entities that can be described as more or less complex. In more complex entities, such as organised chemical, biological, and social compounds, compound-level causal powers are commonly analysed as derivative, so the causal efficacy or causal relevance of the compound

---

[38] *Ivi*: 88.
[39] Taylor 2015: 741.
[40] Yates 2016: 810.

depends upon (i.e. can be attributed *in virtue of*) the causal powers bestowed by the less complex components, which *genuinely* bestow them. However, the question arising at this point – which is a variation of the question "do emergent properties exist?" – is whether it is possible that the complex entity has the causal powers it has *not* in virtue of the causal powers bestowed by its components, but in virtue of something else; in other words, the question is whether it is possible for the complex entity at issue to have novel, fundamental, non-derived, *emergent* causal powers – it being understood that these novel causal powers could influence the causal powers had by the components. Eventually and more generally, are there cases of high-level causal powers that do not directly derive from *other* more fundamental, lower-level causal powers?

David Yates gave an affirmative answer to this question, noting that while it is true that higher level properties are sometimes identified by the causal powers of their lower level realizers (and in that case the higher properties can be reduced to the lower ones), it is also true that other times they are identified by "non-causal" features such as, for example, some structural features. Yates has therefore introduced the distinction between two types of realization which he has defined as *functional* and *qualitative*. David Yates's distinction between functional and qualitative realisation reflects these questions. While in functional realisation "the powers of the realized property are derived from those of its realizers",[41] in qualitative realisation the "realizers do not realize it [the realized property] by bestowing causal powers",[42] but rather by meeting certain non-causal specifications. Yates provides an example of qualitative realisation focusing on some causally relevant properties of the molecule of water that do not merely descend from the causal powers of the atoms composing it. As already seen in the previous paragraph, the H_2O molecule has certain causal powers in virtue of a non-causal feature, i.e. its spatial or geometric structure. In this case, therefore, the higher-level entity exhibits some causal powers that are not grounded in further causal powers, like a traditional view of causation would suggest, but in something else. Yates states, eventually, that those causal powers are novel, fundamental emergent powers bestowed by the emergent property *having* [a particular] *molecular geometry*, which partially determines the higher-level powers alongside with the individual properties of the components. Here we are facing a *joint* determination because while "it's impossible to isolate a token effect of which G [*molecular geometry*] alone is the cause", it's also true that

[41] Yates 2016: 818.
[42] Ibidem.

the higher-level powers are not even derived from the basic physical components alone. In this framework, higher-level compounds are both realized – even if qualitatively and not functionally – and strongly emergent – because causally efficacious: the powers they exert do not derive from the lower level components, for they are novel conditional powers realised by non-causal structures. This circumstance suggests, eventually, that physically acceptable emergent properties and powers are not just possible, but actually real, and what should be abandoned to accommodate these emergent phenomena into our ontology is not physical acceptability but functional realisation together with the classical account of causation based on exerting *individual* powers. Qualitative realisation and an extended view of causation, therefore, may change the cards on the table and open the debate to new possibilities.

As we saw in Chapter I, Carl Gillett's suggestion is similar in providing a revision of the notion of causal determination in the framework of emergence. As we saw in the first chapter, Gillett highlights that causal, "productive" contribution is not the only way in which a property can "make a difference" in the world. Differently from reductionists, Gillett states that there are three ways in which this could happen, while reductionism just considers the first one:

> *Either* (i) the property instance solely *contributes powers* [emphasis mine] to individuals at t; *or* (ii) the property instance solely determines the powers contributed to individuals by other property instances at t; *or* (iii) at time t the property instance both contributes powers to individuals itself and also determines the powers contributed to individuals by other property instances.[43]

While reductionists adopt the view for which causation implies the instantiation of powers and activities, scientific emergentists, in Gillett's opinion, admit an additional form of "non-causal" determination – and the kind of determination that composed or macro-entities exert on their components (*machresis*) belongs to this last kind. On the one hand, "non-causal" determination is *role-shaping* or *role-constraining*. It does not imply the exertion of powers or activities, but just the formation of certain structural or relational conditions able to determine the behaviours of the components in a way which, in its own turn, determines the behaviour of the compound. On the other hand, causal production is *role-filling* in providing

[43] Gillett 2016: 209.

the appropriate powers to fill the roles required to make an effect. These two different kinds of determination are both able to make differences in the world, even if usually just the latter is properly considered as involving causal processes and producing effective real changes. Now, in his 2016 book Gillett defines machresis as "non-causal", for machresis does not involve activities and causation requires activities. However, Gillet later admits that the definition was not particularly fortunate[44] and created confusion. In our opinion, to avoid misunderstandings a general analysis and reconsideration of causation is required. Given a more liberal view of causal efficacy involving (i) functionally realized individual power, (ii) qualitatively realized extrinsic powers, and (iii) determinative constraints, emergent properties seem legitimately efficacious, and this conclusion is coherent with the claims of all those philosophers and scientists who highlight how ignoring the causal relevance of complex structures in the explanation and comprehension of nature is both an unsuitable methodological stance and an unreasoning metaphysical attitude.

Before further developing the analysis of the extended model of causation we have just mentioned, I would like to make some preliminary remarks about the fact that in the literature there is a similar model called the *difference-making account of causation*. This model suggests that causation is equal to difference-making, which is something also suggested by the expression often used by Gillett in defining determinative properties or relations as able to "make a difference" to the powers of individuals. The notion of difference, moreover, is relevant in the field of complexity as well, as information is a structural or spatial difference that can make a difference at higher levels. However, the *difference-making* account of causation found in literature has some features that do not make it useful to clarify the debate on emergence. That is why I specified that the model I would like to propose here is *inspired* by this difference-making model, without however coinciding with it. As we shall see shortly, the canonical model is metaphysically neutral, while the view I would like to recommend has relevant metaphysical implications and can help to understand when and why complex entities that appear to manifest emergent properties and powers are actually efficacious from a causal/determinative point of view. Let us therefore see how this difference-making model is conceived in the literature.

[44] Personal communication.

A concise examination of the debate about the relationships between more classical accounts of causation and this alternative account has been already articulated by Alyssa Ney.[45] Ney outlines the two accounts in the following terms:

> First, there are those who seek out *a physical account of causation* [emphasis mine]. The project in this case is largely empirical. Look to our fundamental, scientific theories and attempt to discover those features that might characterize all actual, causal relations. […] In contrast, and more commonly these days, there are those who pursue *difference-making accounts of causation* [emphasis mine]. […] In general, the aim is to provide an account that captures the truth or assertability of most of the causal claims we make in ordinary circumstances.[46]

Physical accounts of causation were the direct answer to Hume's view of causation as a connection created by the mind but not really present in reality. These answers state that causation, on the contrary, implies a physical connection between the causes and their effects, and this connection is intended in terms of physical processes such as energy flow or transference[47]. Usually, as we saw in Ney's words, these accounts of causation are opposed to an arrey of difference-making accounts of causation which can be defined in different ways. Ney says that these accounts take usually two shapes. The first is a counter-factual account, for which, given a cause *c* and an effect *e*: "(i) c and e both actually occur. (ii) If c had not occurred, then e would not have occurred"[48]. This idea can be found in Peter Menzies, for instance, who defines the truth conditions for causal relevance in the following terms:

> The state S1 makes a difference to the state S2 in the actual world just in case (i) if in any relevantly similar possible situation S1 holds, S2 also holds; and (ii) if in any relevantly similar situation world S1 does not hold, S2 does not hold.[49]

The second meaning difference-making can assume is probabilistic. In this account *c* causes *e* when the occurrence of *c* raises the probabilities of the occurrence of *e*. In

[45] Ney 2009.
[46] *Ivi*: 738.
[47] Some authors who assume this view of causation are Wesley Salmon and Phil Dowe. See Schaffer 2004.
[48] Ney 2009: 738.
[49] Menzies 2013: 73.

Hickcock's words "the probability that E occurs, given that C occurs, is higher than the unconditional probability that E occurs"[50].

Now, the counter-factual and the probabilistic accounts of causation are clearly metaphysically neutral and, accordingly, Ney points out that the physical account of causation is not incompatible with the two versions of difference-making account. While the former is based on science, the latter tries to elucidate more superficial and commonsensical attributions of causation, but there is no reason to considered them as opposite views describing in conflicting terms the very same phenomena.

In our framework, the ability of making differences is something relevant but cannot be just defined in modal terms as happens in the canonical difference-making accounts. In our extended view of causation, the differences produced by emergent entities do have a metaphysical specification corresponding to the acceptance of both causal processes involving functionally realized powers and physical activities, and additional phenomena such as *machresis*, qualitative realisation, and other versions of causal determination such as the constraints account characterising complex systems. After analysing this last account, we will eventually provide some reasons for which these kinds of determinative efficacy should be viewed as a kind of causal efficacy. This reason has to do with the necessity to widen the scope of the notion of causality, for the traditional account based on the mere exertion of individual powers cannot always explain how nature works. One can assume that nature works through mechanisms that are not causal and this may be the reason why causation cannot explain everything – this could be obviously argued. In my opinion, however, it is more reasonable to assume that causation is wider than supposed by the traditional view. As stated by Schaffer, words mean what we choose them to mean, and in the case of causation, causation can mean just physical connection, but this is a choice, and this choice is onerous.

4.4.1 *Constraints*

The notion of constraint is not univocal, but has different meanings depending on the context. In mechanics, for example, a constraint is a parameter that restricts the motion possibilities of a body. This means that when a system is constrained, its parts are not free

[50] Hitchcock 2018: § 2.1 Probability-raising and Conditional Probability.

to behave as they do in isolation – where this last condition depends upon the fundamental physical forces alone:

> Typically, the forces are the basic interaction forces operating among basic system components, e.g. gravitational or electromagnetic forces among particles or chemical forces among molecules. If this exhausts the system, its behaviour is said to be free, not under constraint, that is, it has available all of its degrees of freedom, = all of the mutually independent kinds of basic motion its interaction structure permits (translations, rotations, etc.). Constraints then represent *additional forces* [emphasis mine] on the system that also contribute to shaping the system dynamical behaviour […].[51]

A classic example is the one involving a sphere that rolls along an inclined plane. The sphere is not free to move in any direction, but must move along the plane. Its motion is therefore constrained by the external surface on which it rolls. Another example involves a carriage that cannot deviate from the trajectory of the rail on which it runs. Also in this case, if it were subjected only to the fundamental forces, the object would have greater possibilities of movement, but being constrained by the rail it is forced to move in certain ways (and its degrees of freedom decrease). The physical constraints are typically external, in the sense that an object is constrained when it is forced to move due to another external object (the plane, the rail or similar). The notion of constraint referred to in biology, instead, presents different characteristics because the constraint, in biological systems, is not external to the system, but internal and is produced by the same dynamic that it constrains[52].

The notion of constraint we are referring to traditionally traced back to the years between the sixties and the seventies, to two papers written by Michael Polanyi[53] and Howard Hunt Pattee.[54] These authors focused on the fact that certain complex systems cannot be exhaustively described in merely physical or physical-chemical terms given the various conditioning they are subjected to. In Polanyi's view, each organism is subject to two different principles: on the one hand, to its biological higher-level structure, and, on the other hand, to the laws of physics and chemistry. Moreover, the former constrains the latter as follows: "[…] its structure serves as a boundary condition harnessing the physical-

[51] Hooker 2013: 757-8.
[52] See Bich and Mossio 2011.
[53] Polanyi 1968.
[54] Pattee 1970.

chemical processes by which its organs perform their functions. Thus, this system may be called a system under dual control".[55] Polanyi provides an example centred on the nature of a spoken literary composition that can be viewed as formed by five hierarchical levels: the voice, the words, the sentences, the style, and the text. Each level is subject to both the principles of its own level and those of the next higher level: "The voice you produce is shaped into words by a vocabulary; a given vocabulary is shaped into sentences in accordance with a grammar; and the sentences are fitted into a style, which in turn is made to convey the ideas of the composition".[56] The derivation and explanation of the nature of each of these levels, by contrast, follow the opposite downward direction: "You cannot derive a vocabulary from phonetics; you cannot derive grammar from a vocabulary; a correct use of grammar does not account for good style; and a good style does not supply the content of a piece of prose".[57] Biological organisms follow this determinative schema as well, but are formed by many more levels. While the functions and laws belonging to each level are always preserved, additional constraints represented by the higher controlling levels appear and shape the behaviour of the components in order to achieve the purposes of the whole system. Pattee's formulation of this problem is similar. He highlights that while physical laws are "inexorable and incorporeal",[58] constraints are always embodied in material structures such as molecules or membranes. These structures are made of matter, so the fundamental laws of nature are always in place, but at the same time their possible manifestations are constrained. Constraints, therefore, determine the range of the physically possible states becoming actual, which means that systemic constraints reduce the degree of freedom of the components, determining, by so doing, their behaviour. The activity of the constraints can therefore be understood as a sort of negative causality that acts by collapsing the set of possible states of the constrained entities and their respective behaviours to a smaller subset[59] composed only of the states that contribute to the functioning of the higher level system. This negative and constricting activity, however, also corresponds to an enabling capacity, since the constrained system instantiates properties and exerts a determinativity that the components could not manifest if aggregated in an unconstrained way, and in the case of biological systems this type of determinativity and the possibilities that accompanied it have determined the evolution of

---

[55] Polanyi 1968: 1310.
[56] *Ivi*: 1311.
[57] Ibidem.
[58] Pattee 1970: 250.
[59] See Blachowicz 2013 and Hooker 2011.

complexity. It is worth noticing that negative causes are another case of causes not admitted by the view which intends causation in terms of physical processes or exchange of energy, given that in negative causation causes are not physically connected to effects: “Rather, what is causally salient here is the *absence* of a physical connection” [60].

Now, at the end of this chapter I will offer an example of a biological system that manifests macro-constraints: I will examine the characteristics of ant colonies. To anticipate some of the conclusions, however, I can say this: ants, individually taken, can behave in many ways, but when they live in a colony, the set of their *possible* behaviours is limited by the social structure of the colony to the subset of behaviours that are *useful* (to the colony). This is how, put in very few words, high-level structures constrain and determine the states and properties of low-level components. Now, ant nests, like eusocial insect colonies in general, are classic examples of complex systems and in fact the notion of constraint, as well as that of diminishing degrees of freedom, is often used by scientists of complexity to describe the behaviour of these systems. The notion of complexity, in my opinion, is fundamental in the debate on emergence. While it is true that the criteria for emergence have always underlined irriducibily and novelty, it is equally true that emergence has always characterized the study of complex systems. So what is the connection between irreducibility, novelty and complexity? Does emergence only affect complex systems? The answer to this last question, as we can anticipate, is obviously negative. Since the beginning of this dissertation I have stressed how emergent phenomena are heterogeneous and affect different domains of reality. We can build several models of emergence reflecting the presence of different manifestations of the phenomenon, but they present a set of relevant properties that lead us to place them in the same ontological category, i.e. the so-called "emergent phenomena". The emergence that complex systems exhibit, therefore, is not the *real* or *authentic* one, since there is no competition for this title, but, in my opinion, it is particularly representative and worth focusing on. As we will see in the conclusions, the idea that emergence can be defined through a cluster of properties is accompanied by the idea that depending on different ontological areas, there are different types of emergence which are typical of those particular regions. However, there are kinds of emergence that are more interesting than others, given that the interestingness is obviously dependent upon the particular interest of the observer. In my case, I think that emergence can be instructive when it is metaphysical rather than just epistemological,

[60] See Schaffer 2004: 204.

when it involvs two kinds of causal determination, and when it is widespread. First, metaphysical emergence is informative of the metaphysical structure of the world, so it is intresting if one wants to better understand the way in which reality is organised. Second, and related to the first point, emergence can be determinative in two different ways (at least): on the one hand, it can constrain its parts in a top-down manner, and, on the other hand, it can be "horizontally" determinative at the high-level, and understanding the nature of these two forms of causal determination can contribute to the understanding of the structure of the world. Third, and related to the first and second points, if emergence is widespread, it means that this kind of organisation is not just limited to special parts of our ontology, but is a common trait of reality.

Now, the kind of emergence observed in complex systems manifest these three features, so I will consider it more representative than others: complexity concerns a great number of phenomena that can be found at many different levels of realty and emergent complex phenomena exhibit what I think are the most interesting properties associated with emergence. For this reason, I dedicate the rest of the chapter to complexity, first defining complex systems, then outlining how they exhibit emergence, and finally providing an example concerning ants and a model delineating the relazionships involved in complex systems exhibiting emergence. If irreducibility and novelty are widely considered as criteria of emergence and if irreducibility is often due to the presence of novelty, the latter is certainly made possible by the way reality is organised, i.e. in forms that are rarely simple, additive and linear. Let's talk, therefore, about complexity.

4.4.2 *Complex Systems*

A complex system is a set of elements that once together manifest sophisticated and organised behaviour. Examples of complex systems are not only social insect colonies, as we anticipated, but also the brain, the Web, or economies. They are highly various phenomena which at the right degree of abstraction exhibit common properties.

There are mainly two approaches which could be adopted to define a complex system, and their integration is essential to a deep understanding of the phenomena. The first, which is an ontological approach, focuses on the intrinsic organisation of complex systems and tries to describe their properties, setting the conditions suitable to distinguish complex systems from merely *complicated*[61] ones. The second approach, which is epistemological, underlines the immense difficulties that emerge in understanding, explaining, modelling, and creating complex systems, and makes these difficulties the mark of complexity[62].

Despite these two strategies and the abundant studies dedicated to complex systems in the past decades, a precise definition of complexity is far to be obtained, as the history of the concept testifies.[63] One reason for this situation may be that complexity, as well as emergence, can be identified in many different scientific disciplines, which, therefore, need different definitions, often mutually incompatible. In other words, no single definition could be exhaustive in describing all phenomena which intuitively exhibit complexity. However, this confusion should not be discouraging, as there are several cases of useful, although unclear, scientific concepts. Molecular biology, for instance, is today one of the most promising scientific fields, but what a gene is, is still deeply problematic. Same for astrophysics: we know that the universe is 95% composed of dark matter and dark energy, and we have no clear ideas about what they are, but this lacuna does not prevent physicists from formulating theories about the origin, the evolution, and the actual state of the universe.

The root of the word *complexity* is the Latin verb *plectere* which means *twine* or *blend*, and, in fact, complex systems could be described by this idea: a complex system is a network (or a network of networks) of many simple components intertwined through nonlinearity interactions. Although the system lacks central control, it presents, among others, two interesting features. Firstly, it results able to both use and produce information. Secondly, it exhibit the formation of sophisticated structures and gives rise to global patterns of behaviours that its components do not show. Eventually, the systemic structures are able to influence and determine the states of the components by limiting their possible states to a sub-set of states functional to the whole. In the framework of the science of complexity, a way to better comprehend the nature of these properties has been to focus on the notions of self-organisation and emergence. It could be said, in other terms, that the many simple components of the system spontaneously

[61] Holland 2014.
[62] In 2001, the physicist Seth Lloyd wrote a paper about the (many) different ways to measure complexity. Different disciplines deal with complexity, and so different measurements have been formulated. Nonetheless, the debate gravitates around few questions: “1. How hard is it [a complex system] to describe? 2. How hard is it to create? 3. What is its degree of organization?” (Lloyd 2001:7).
[63] The above cited paper of Seth Lloyd (Lloyd 2007) lists forty possible ways of measuring complexity, related to different disciplines which deal with these feature.

self-organise in a way that global, higher scale patterns *emerge* from their interactions, and new properties and causal/determinative structures appear, such as the sort of downward determination we mentioned before. These emerging properties, which are called *systemic,* are the properties that the system instantiates *as a whole*, and this represented one way to describe the meaning of the often repeated holistic quote: "the whole is more than the sum of its parts".[64]

The fact that properties of the parts are not additive produces a significant inability to predict a large number of phenomena that the reductionist programme is unable to explain. A short but significant list was provided by Melanie Mitchell, Professor of Computer Science at Portland State University: the climate, the adaptive capacity of living organisms, the diseases that afflict them, the economic, political and cultural dynamics affecting our societies, the growth and effects of technology, the nature of intelligence and the ability to artificially recreate it.[65]

Now, although it is impossible to provide an exhaustive definition of the concept of complexity, in the following paragraphs I will outline the most relevant characteristics of complex systems by focusing mainly on three issues: their non-linearity, their ability to process information and their ability to produce emergent behavioural patterns that exhibit causal-determinative effectiveness on the parts and other systems at the same level. As we shall see, in the study of complex systems the notion of constraint is fundamental and presents relevant similarities with David Yates' qualitative realization model and with Carl Gillett's macretic determination model The next paragraphs are dedicated to an analysis of the kind of information processing characterising complex systems and the self-organisation patterns emerging from these processes. As we will see, in the study of complex systems, the notion of constraint is fundamental and presents relevant similarities with David Yates' model of qualitative realisation and Carl Gillett's account of machretic determination.

4.4.3 *Information*

The ability to self-organise is one of the most obscure features of complex systems, and a promising way of approaching this feature is by exploiting the notion of

[64] The concept, expressed in similar words, is present both in Aristotle's *Metaphysics* (Book VIII, 1045a.8–10), and in Euclid's *Elements* (Book I, Common Notion 5).
[65] Mitchell 2009: x.

information processing. A complex system is a system which encodes, stores and uses a specific amount of information: more than a simple system, but less than a chaotic system, and this specific quantity seems connected to the possibility of self-organisation. At this point, the relevant questions are how a complex system stores, structures and uses information, and how it produces it, in order to self-organise and pursue its systemic goals. Let's start with modern definitions.

In 1948, Claude Shannon, an electrical engineer and mathematician working at the Bell Labs, published his seminal paper *A Mathematical Theory of Communication* in the Bell System Technical Journal, and the next year, when the importance of Shannon's theory for science in general became clear, the paper was republished in a book containing another writing – a kind of commentary to Shannon's one – by Warren Weaver.[66] Weaver, an American mathematician who would have long collaborated with Shannon, wrote that, as far as communication is concerned, three kinds of problems emerge at different levels:

> LEVEL A. How accurately can the symbols of communication be transmitted? (The technical problem.)
>
> LEVEL B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)
>
> LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.).[67]

We do not focus on the third problem for a matter of pertinence and consider the first and the second one, which reflect two different ways of interpreting information.

In the first case, that of the "technical" level, the approach to information is syntactical. Information is a sequence of symbols transmitted from a source to a receiver, and the meaning of the symbols is irrelevant. In this case, information, as its Latin etymology suggests, is only a defined *form* which could or could not convey content, and this kind of information, as well as the problems it entails (e.g. data compression, or signal processing), is the one studied by Shannon's information theory.

[66] Shannon & Weaver 1999.
[67] *Ivi*: 4.

In the second case, the approach is semantic: information is connected with its content, or, in other words, with its *aboutness* and representation. Semantic information expresses, therefore, a meaning, has a reference in the world, and could be true or false.

As highlighted by Luciano Floridi, information, as complexity, "is notoriously a polymorphic phenomenon and a polysemantic concept so, as an explicandum, it can be associated with several explanations".[68] In recent years, many scientific fields have adopted a general definition of information (GDI) to reduce this ambiguity. GDI states that something is an instance of information if: i) it consists of one or more data, ii) the data are well-structured, and iii) the well-structured data have a meaning.[69]

Now, although the importance of meaning, we think we could start from the first level described by Warren, whose problems have been tackled by Shannon's theory. Shannon formulates a quantitative and probabilistic theory of information, in which meaning is irrelevant. What is relevant is the data that physically constitute the message and the probability for that data to be sent from the source of information to the receiver. As Shannon states: "The significant aspect [of communication] is that the actual message is one selected from a set of possible messages".[70]

For Shannon, therefore, information means data communication and the difference between two information is in that the *form*[71] of the message is different from the *form* of other possible messages.

This spatial, physical difference, which is merely syntactical, is what characterises information. It is worth noting that syntax, in this context, should be considered in a broad sense: it is "what determines the form, construction, composition, or structuring of something".[72] In other words, syntax defines the *physical architecture* of data, and this difference in architecture recalls the famous "difference that makes a difference",[73] an often mentioned quote by Gregory Bateson.

Focusing on the idea of difference is fundamental. The unit of information, the *bit* ("binary digit"), is the representation, through a binary code (0;1) of the physical difference between two possible states, such as ON and OFF, or heads and tails. Systems such as a light switch or a coin, in other words, present physically different possible

[68] Floridi 2014: 81.
[69] See Floridi 2017: §1.2 The data-based definition of information.
[70] Shannon 1948: 379.
[71] See Cohen 2006.
[72] Floridi 2014: 84.
[73] Bateson 1972: 465.

states that can be represented by the twofold value of the bit, 0 or 1. Information should be seen, therefore, as *the representation of a physical difference*, and this difference at the physical level makes a difference at the systemic level,[74] as we will briefly see. Complementary, any physical system which presents different states already encodes information, and the more possible states the system has, the more information the system encodes. To give an illustration of this, just imagine a "unary device", like the famous Raven described by Edgar Allan Poe[75] who answers "Nevermore", regardless of the question asked. The Raven is a unary device for it is a system that provides the very same output for any incoming input. Unary devices have only one possible state, so the amount of information provided by such systems is equal to zero, as the outcome has a probability of 1 (100%). Differently put, Poe's Raven answer does not provide information because it cannot increase knowledge, or, as states Luciano Floridi: "we already know the outcome of the communication exchange, so our ignorance (expressed by our question) cannot be decreased".[76] By contrast, a two-state system, like the coin mentioned above, offers more information because for any coin toss we have two different possible outcomes, with the same probability of 0.5 (50%). In this case, the answer to our virtual question is unknown until the discovery of the outcome, when a certain amount[77] of information is produced. Suppose, finally, to have a six-faced die. The die system encodes even more information than a coin, because the number of its possible states is six, and in fact we need more than two bits to measure its information outcome. From Poe's Raven (one state), we passed to coins (two states) and to dices (six states), but it is possible to continue until biological systems such as the DNA, which encodes more than 6 billion of information, with a terrific corresponding quantity of possible states (genetic configurations), each of which makes a big difference at the level of the organism. It is not a causality that the notion of information is so central in

---

[74] As already, about the legitimacy of talking about levels (a topic which recalls a metaphysical layered view of reality) we will widely discuss in Chapter II.
[75] Poe 1903: 125-128
[76] See Floridi, 2007: § 2.1 The mathematical theory of communication
[77] In particular, the quantity of information produced is proportional to the number of possible states of the system. If we have one bit of information (one coin), the total number of possible states N will be two, while if we have two bits of information (two coins), N will be equal to four, and so on. Conversely, if we have a system with eight possible states (N equal to eight), we will have three bits of information, and so on. It is important to note, here, that the proportionality at issue is exponential (i.e. nonlinear). If one bit of information represents only two states (N=2), a handful of bits, like twenty or thirty bits, represent millions or billions of possible states. This proportionality follows the simple mathematical function: N=2n, where *n* in exponent is the number of bits. With *n* equal to 300, for instance, we approximately reach the number of particles in the universe.

biology. As stated by Eirs Szathmary and John Maynard Smith, "developmental biology can be seen as the study of how information in the genome is translated into adult structure, and evolutionary biology of how the information came to be there in the first place".[78]

Now, as far as the questions formulated at the beginning of the paragraph are concerned – how complex systems store and manage information, and how a complex system produces new information from the acquired one – some answers can be provided. Complex systems store and encode a huge quantity of information in having a massive amount of possible states and configurations, more than simple systems, less than chaotic systems. This last point, however, entails that the mere increase of the quantity of information in the system does not straightforwardly correspond to its complexification. Complexity, rather than depending on the quantity of available information, depends upon its organisation, i.e. on the kind of possible configurations the system components assume.

4.4.4 *Self-organisation*

Quoting Bateson, we mentioned "a difference which makes a difference". What we mean is that the difference at the physical level – i.e. the distinction between different physical states – corresponds to information that in dynamical interacting systems produces both local and systemic consequences.

While information is generated by differences, however, information processing emerges from interaction.[79] Two systems interact when the state of the first is influenced by, or dependent upon, the state of the second, and this interaction generates new further possible states and, therefore, novel information. Let me provide a couple of examples. First, suppose to have a coin and to flip it fifty times. It is clear that no coin toss will be influenced by the previous one. Heads and tails would still have the same outcome probability, regardless the coin toss is the first, the tenth, or the forty-ninth. In statistics, this feature is described by the idea of *variables uncorrelation*, for which two random variables are not correlated if their dependence and their covariance are equal to zero. By contrast, variables that are correlated present a nonlinear relationship, and they

[78] Szathmary and Maynard Smith 1995: 231.
[79] Lizier 2013: 13

influence each other. If the outcome of a flipping coin well represents the first kind of variables, as far as the second type is concerned, we could take as an illustration an everyday, useful object like the thermostat.

A thermostat is a control device. It measures the temperature of the environment and turns on or off a heating system to maintain a steady temperature state, as close as possible to the fixed set point (e.g. 21°C). In this case, we have the first variable which is the external temperature, and the second variable which is the output of the sensor which measures the temperature at issue. If the temperature is below a certain threshold, the thermostat will not turn on the heating device, while if it is above it, the thermostat will turn it on. The outcome of the system (on/off) is, therefore, dependent on the input (below/above the threshold), but the input itself (the temperature of the room) is dependent on the output. The heating system, indeed, will increase or decrease the external temperature, which will be measured again, and will influence the outcome of the thermostat and so on and so forth.

A system like a thermostat presents circular processes called feedback loops. They are self-regulating control processes that "connect output signals back to their inputs".[80] Feedback loops are essential in biological systems: homeostasis, for instance, is based on positive and negative feedback loops, and it is what allows the cells to be flexible, functional and robust.[81] In these cases, variables are dependent on each other, and their relationship is a covariance one: the activation of the heating device depends on external temperature, and external temperature depends on the activation of the heating device. As said by the computational neuroscientist Péter Érdi: "In a feedback system, there is no clear discrimination between "causes" and "effects", since the output influences the input".[82]

To summarise, systems with correlated variables present dependence and interaction between their components, while systems with uncorrelated variables do not. In information and probability theory, this circumstance is measured by a quantity called mutual information (MI). MI measures the amount of information that one variable can convey about the other. In other words, it is a measure of mutual dependence. In systems with dependent (i.e. correlated) variables, like the thermostat, it is possible to obtain information about one variable through the other. By contrast, in systems with

[80] Brandman & Mayer 2008: 390
[81] See Érdi 2008: 32-33
[82] Érdi 2008: 8.

uncorrelated variables, like a flipping coin, the knowledge of the value of one of them (e.g. the outcome of one toss) provides no information about the other. The mutual information of a pair of coin tosses is therefore equal to zero, while the input and the output of a thermostat carry a significant amount of it.

Mutual information is an essential notion for the sciences of complexity, so much so that over the last fifty years it has been consistently pointed out as a promising measure of complexity.[83] Now, beyond the problem of the quantification of complexity - which is an interesting issue but tangential to the purposes of this work - it is worth highlighting the philosophical meaning of mutual information: it is possible to obtain information on variable *a* knowing the value of variable *b* when the entities that *a* and *b* describe (the parts of the system) are inserted in a strong relational structure within which they regularly interact. At this point, the question that arises concerns the nature of these relations. What kind of interactions are they? And what do they involve? It is at this point that the notion of self-organization comes into play.

By self-organisation, it is usually[84] intended "a phenomenon under which a dynamical system exhibits the tendency to create organisation "out of it-self", without being driven by an external system, in particular, not in a "top-down" way".[85] Another definition is provided by Camazine *et al.*: "Self-organization is a process in which pattern at the global level of a system emerges solely from numerous interactions among the lower-level components of the system. Moreover, the rules specifying interactions among the system's components are executed using only local information, without reference to the global pattern".[86]

A self-organised system manifests, in other words, three critical features:

1. It has a structural and/or functional sophisticated nature.
2. It has neither internal, nor external, central control.
3. Global high-level patterns are produced by local lower-level dynamics which use local information alone.

As far as this last point is concerned, an instance can be birds' flocking behaviour. In a flock of flying birds, animals follow few simple rules well described by the simulation model *Boids* developed by Craig Reynolds[87] in 1987. The model exhibits the

---

[83] See Emden 1971, Chaitin 1979, Grassberger 1986, Liezer 2013, Kanwal, Grochow, & Ay 2017.
[84] See Lizier 2013: 16.
[85] Polani, 2008: 25.
[86] Camazine et al. 2001.
[87] Reynolds 1987.

importance of three steering behaviours. The first is *cohesion*: the birds steer towards the centre of the flock. Second, *separation*: the birds maintain an average distance from each other. Last, *alignment*: the birds follow the average flight trajectory defined by the neighbours. Notably, an essential condition of this model is that birds react *only* to immediate neighbours, which means that no bird has global information about the flock as a whole. The global pattern of flocking, therefore, emerges by *local* information processing. Actually, the same applies for schools of fish or social insect colonies, although the collective rules are different. Furthermore, it is worth noting that no central control organises local dynamics in these biological systems. As written by Mitchel Resnick in his *Turtles, Termites and Traffic Jams*: "The flock is organized without an organizer, coordinated without a coordinator".[88] In sum, it could be said that the behaviour of the single animal, rather than responding to an external command (nor to an internal programming), is a function of the actions of its neighbours, as happens in Cellular Automata, where at each temporal step t_1, $t_2 \dots t_n$, every cell composing the model synchronously updates its state as a function of neighbours' previous states.

Now, given this brief description of the most relevant features of complex systems – the huge numbers of components, the nonlinear, mutually dependent relationships between the parts, and the presence of global patterns of behaviour arising from local computation – let's see an example that can be useful to understand the kind of determination the system exerts on the components.

4.4.5 *Ant colonies*

We stated that complex systems exhibit global patterns of behaviour from local computation. Examples are the birds composing a flock, as well as the Cellular Automata cells, which change their behaviour or "update" as a function of neighbourhood. As a consequence of these local computations, eventually, complex patterns emerge at the system level.

Social insect colonies are a fascinating example of this dynamics. As written by the biologist Deborah Gordon, who has been studying ants for almost thirty years: "Thinking about collective behaviour generates sticky questions about the relation between behaviour at the level of the individual participants and the level of the group

[88] Resnick 1997: 3.

or system".[89] In other words, the main problem related to collective behaviour is how it is possible that interactions among individual ants produce global, predictable and goal-directed outcomes.

First of all, let me provide a brief description of ant colonies.

Biologists identified approximately 12.000 species of ants, and places like the tropics hide many other species not yet studied. Ants evolved by wasps in the Cretaceous, 140 million years ago. They are present in all continents, except Antarctica, and their size, colour, diet and behaviour vary in response to the different environments in which they live. Among all the known species, however, only 50 have been adequately studied, so our knowledge about ants is significantly incomplete. Nonetheless, we know that all species of ants share a significant feature: they all live in colonies. Colonies could be of different dimensions, from little nests with a bunch of ants to supercolonies composed by sub-colonies containing several millions of individuals. In every colony there is one, or more than one, reproductive female – the queen – and many sterile working ants that perform different tasks, such as hunting and searching for food (foraging), defending and maintaining the nest, taking care of the queen, of the eggs and of the larvae, and so on.

Although this general distinction between queen and workers, however, individual ants' behaviour is pretty unpredictable, as ants do not seem genetically "programmed" to perform a task, as some studies suggested[90]. Many theories correlate size and task, suggesting that particular tasks are better suited to ants of a particular size. Nonetheless, only a few genera of ants exhibit size variations, merely 44 of 263.[91] Furthermore, ants' tasks are not rigidly determined. Ants switch their task in case of necessity, namely when colony's needs vary in response to environmental changes. For instance: "An animal steps on the nest, or rain seeps in, and nest repairs are needed. There is a windfall of food, or there is a shortage. The changing environment continually shifts the numbers of ants required to perform each task, to repair the nest or collect food".[92] Moreover, the so-called allocation task, which is the process which tunes the right number of workers needed in a particular situation, also depends on colony dimension[93] and ants' age: "A worker follows a predictable pattern of changes between tasks as it ages. This pattern

[89] Gordon 2016: 514.
[90] See Oster & Wilson 1978, Trible & Kronauer 2017.
[91] Gordon 2010: 30.
[92] Gordon 2010: 24.
[93] Detrain & Deneubour 2006: § 3.4 *Population size matters*.

usually leads from workers first doing a safe task like brood care, to ending their lives doing the most dangerous tasks like foraging”.[94]

In sum, what an ant does is not determined by its individual, genetic identity, but by the network of interactions between it, the colony, and the external environment. Ants behaviour, therefore, depend on a relational structure, and, as we will see, these relationships are nonlinear, consist of local information processing, and produce higher-level behavioural patterns.

First, ant interactions create a nonlinear system. Ants interact through olfaction of chemicals or mechanical grasping of legs and antennae. This communication system permits them to share information through feedback mechanisms. To give an illustration, imagine a forager searching for some food outside the nest. Notably, individual search behaviour is chaotic,[95] as the ant does not know where the food is, and so a random walking could drive it to new paths and discoveries. Eventually, the ant finds a source of food and brings some of it back to the colony, leaving on his return journey a trail of pheromones. Once arrived at the nest, it recruits other ants which, following the path marked by chemicals, will reach the food source as well. While the search behaviour of the first ant was random, other ants’ behaviour is driven by nestmate’s chemicals, making the emergent global behaviour of the ants something completely different from the sum of the behaviours of the individuals taken in isolation.[96]

However, as recent studies illustrate[97] also the reaction of ants to pheromone trail is not linear. Ants react to them as a function of chemical concentration, but notwithstanding to the strength of this concentration, ants could lose trails and move unpredictably. To put it in another way, ant behaviour is flexible. It is never entirely determined by chemicals: “the ants ‘choose’ between possible solutions”.[98]

Eventually, ant colonies are nonlinear complex systems for another reason. Every ant is different from others and therefore its behaviour would be idiosyncratic and dependent on its past: “the genetic background, the caste belonging, the sensory integration or the own experience of each worker determine its likelihood to interact

---

[94] Tripet & Nonacs 2004: 863.
[95] See Cole 1991, Solé, Miramontes & Goodwin 1993, Sumpter & Beekman 2002.
[96] See Solé, Miramontes & Goodwin 1993.
[97] See Sumpter & Beekman 2002, Sumpter, Mann, & Perna 2012, Nicolis, & Deneubourg 1999.
[98] Sumpter & Beekman 2002: 274.

with nestmates, to react to environmental cues or to respond to signals emitted by congeners”.[99]

In sum, ants’ interactions are nonlinear in virtue of several factors. Among them, we mentioned the stochasticity and individuality of the individual ant and the interrelated communications between ants as a group. There is a last significant detail we have to consider, however, namely the direct, dynamical response of every individual ant to environment changes. Gordon defines as “algorithms”[100] the processes which generate collective behaviour from lower-level interactions and illustrates that these algorithms evolved in response to the environment, like every other phenotypical trait. Many environmental features can indeed influence colonies, for example, food distribution, soil physical-chemical properties, presence of predators and competitors, temperature, rates of signals propagation, and many others.

As stated by Detrain and Deneubourg:

> Looking at ants’ collective patterns as self-organized systems has underlined the role of the environment. Instead of acting simply as a constraint on ants’ behaviour, the environment and its properties now appear as actors in the pattern formation process.[101]

Colonies’ behaviour should, therefore, be approached from an ecological perspective, which takes into account many factors and not only the individual properties of the components of the colony.

Second, ant colonies are an example of local information processing giving rise to global behaviours. As was mentioned, ants react to the environment and to nestmates, which means that their actual behaviours depend on these two factors. An ant can be active or inactive, can work inside or outside the nest, it can search for food, patrols the surrounding of the colony, manages the refuse pile, and so on. What it does depends both on the actual environmental conditions, and on what its proximate mates do. Social insect colonies, in this, are not different from other animals who live in groups, such as schools of fish, flocks of birds, mammals’ herds and so forth. Notably, in all these cases, individuals only react to their neighbourhood. Even in case of sophisticated behaviours,

---

[99] Detrain & Deneubour 2006: 177.
[100] Gordon 2016: 514.
[101] Detrain & Deneubour 2006: 176.

like building or moving the colony, cultivating fungi and yeast, farming and milking aphids, or creating rafts, no ant knows neither the global state of the group nor what the colony is currently achieving. Actually, there is no one at all which is aware of colony's global state (except the observing entomologist, if he is present). Colony's behaviour emerges merely from local information processing, without previous programming nor central driving control, and how it is possible is one of the most difficult problems system biology is confronting.

What this example suggests, eventually, is that the number of an individual ant's possible behaviours is huge, but when the ant is living into a colony, this set of possible behaviours is limited to the sub-set of possible behaviours that are useful to the colony. In other terms: "the state of the system taken as a whole plays a role in determining the allowable states of the system's constituents".[102] With the formation of a colony, therefore, a system-wide, macro-constraint limits the states of individuals and by doing so new properties such as the ones previously mentioned – building a colony, moving it in virtue of the environment, cultivating fungi and yeast, breeding, protecting and milking aphids, creating rafts and so on – appear and make differences at both the individual, systemic, and ecological level. To include these new properties among the causally efficacious ones, however, a further move in the metaphysics of causality might be necessary, and this is the topic of the next paragraph.

4.4.6 *The extended view*

In § 4.3, we said that Kim's Causal Inheritance Principle can apply to those cases in which being causally determinative corresponds to bestowing and exerting powers, and, similarly, we noticed that the traditional view of realisation implies that being realised means being *causally* or *functionally* realised. However, we also saw that in many natural structures, such as in molecules or in complex systems like ant colonies, being determinative and producing efficacious properties is not equal to merely bestowing causal powers. Now, the difference between the cases in which powers are supposed to be directly exerted, and those in which they are not, can be clarified analysing their nature.

In the metaphysics of powers, powers are attributed of some distinctive features that

[102] Lawhead 2015: 7.

seem oddly conflicting with the very possibility of emergence. This could mean that as long as this account of powers (and causation) is taken into consideration, emergence will be *a priori* excluded from our metaphysics. To anticipate the content of this paragraph, we may say that while powers are usually considered *intrinsic* properties of their bearers exerting an *active*, powerful role, emergent properties and powers are often characterised in a way that suggests that they might be defined as *diffused* in structures rather than possessed by particular individuals, and as *constraining* powers, not involving any individual, intrinsic activity. It might be reasonable, therefore, to widen the scope of the notion of causal power to these kinds of power too, so that it will be possible to accommodate the many phenomena involving these alternative forms of determination among the effective players of reality. Let's now focus on the two features traditionally characterising powers: intrinsicality and activity.

As stated by Neil Williams, there is one thesis almost universally accepted among the metaphysicians of powers, namely the *Intrisicality Theses*. The thesis states that powers are intrinsic properties[103], so their possession by an object depends upon that object alone, and identical objects will have identical powers, regardless their context. Being intrinsic, moreover, means being essentially associated with a bearer, so powers are not just intrinsic, but they are also individual. Indeed, talking about powers always involve talking about their bearers or possessors:

> A power is a property that provides its possessor with the ability to bring about some states of affairs—the *manifestations* of the power—when the power finds itself in the appropriate circumstances (themselves states of affairs). Actually, to say it provides its possessor with abilities is misleading: a power *is* that object's ability to bring about those states of affairs.[104]

Powers, therefore, are always possessed by particular objects, and each power can be localised – at least in principle – in those individuals possessing it.

The second feature usually attributed to powers – a feature that is also mentioned in the passage we just read – is their *active* nature. Powers actively dispose their bearers to do something[105] or to bring about some states of affairs – i.e. some manifestations. This is the

[103] See Mumford 1998: 74, Heil 2003: 195, Molnar 2003: 129, Williams in Marmodoro 2013: 85 and Williams 2019: 68.
[104] Williams 2019: 49.
[105] See Lowe in Marmodoro 2013: 10 "[…] a power is always a power *to do something*".

reason why the words “power” and “disposition” are often used interchangeably: powers confer abilities to their possessors and these abilities actively *dispose* the bearers to behave in certain ways:

> The causal powers […] belong in the category of dispositional properties, along with propensities […]. What these active properties all have in common is their dispositionality. The categorical properties, in contrast, are essentially passive; since there is nothing that their bearers are necessarily disposed to do just in virtue of their having these properties.[106]

Now, if being causally efficacious means exerting such kind of individual, active powers, then for many natural, psychological, and social phenomena it will be impossible to be genuinely and not merely derivatively causally efficacious. By definition, basic causal properties are instantiated by basic fundamental entities, but if just basic causal properties are considered genuinely causal, it is obvious that just basic fundamental entities will be causally efficacious. This circumstance, however, might derive by the choice of a too strict account of causal efficacy, rather than by the real inefficacy of all non-basic, more complex phenomena. Structures like molecules, as well as almost every complex system, suggest that in addition to individuals exerting their intrinsic active powers, other kinds of determinative efficacy exist, and the latter are directly responsible for changes in nature and reality as well as the former are.

While powers are usually intended as individual and active, these other forms of causal determination are not directly connected to individuals, because they are, we will say, *equably diffused* in the structures composed by them, e.g. the spatial structure of molecules, or the relational structure of systems such as ant colonies. Properly speaking, in these cases there are no basic, particular individuals exerting intrinsic physical forces. Rather, the determinative power of the whole is *spread* among the components, it is triggered by their relationships, and its *locus* should be localised at the higher level of the structure in a holistic way. Moreover, and this is the second difference, emergent determination is sometimes *constraining* in exerting a sort of negative causal determination, and it is no coincidence that that of constraint is a key concept in those disciplines dealing with complexity and complex systems.

[106] Ellis in Marmodoro 2013: 136.

As we saw in 4.4.1, a *constraint* is something limiting the independent behaviours of the entities subject to it. However, by constraining parts of the system, the structures that exert this kind of coercive influence not only produce negative and disabling effects, but also enable the appearance of new states and behaviours. The structure in which ants are organised, for example, binds their possible behaviours to a particular subset; it acts by pruning the possible states of the component to those that let the system persist[107]. This limitation – which corresponds to the organization of the system – seems to be a mere restriction but just on the components' level since it involves the reduction of their degrees of freedom. This constraint, however, permits the appearance of new structures, new properties and new forces at a systemic level that a free aggregation of those same parts would not have allowed.

Figure 4.1 illustrates these circumstances. High-level complex structures are syncronically composed or realized (blue arrows) by lower-level elements and this fact rules out any dualistic interpretation of emergence: the components are physical and they constitute the higher-level structure. The structure, however, exhibit macro, system-wide, global constraints exerting limitations upon the components (grey arrows). It is possible to visualize the action of the constraints (Fig 4.2) supposing that each component has a set of possible states (the circles around it) that are all viable (blue) when the component is free or in isolation, but are limited to a subset when the component is wrapped into an organized structure. In this last case, the states of the components will be partly consistent with the structure (the blue ones) and partly inconsistent (the white ones), and just the first ones will be admitted by the constraining relationship. When the components are constrained in an organized structure, in other terms, their contribution is conditioned, as Alexander and then Gillett highlighted. If saying that "the whole is more than the sum of its parts" can be misleading, therefore, saying that "parts behave differently when organized in wholes" can be more illuminating. Moreover, constraining relationships have another interesting feature: they are synchronic because consist in negative causation, which is a kind of causation that can be synchronic because no physical processes (which imply a temporal duration) is required. The limiting action of constraints, moreover, entails the emergence of novel forces able to influence the dynamics at the system level (yellow arrows). These novel forces are not basic physical forces possessed by basic physical entities and inherited by the high-level structure: rather, they are novel forces qualitatively realised by the structure itself (in a

[107] Morowitz 2002.

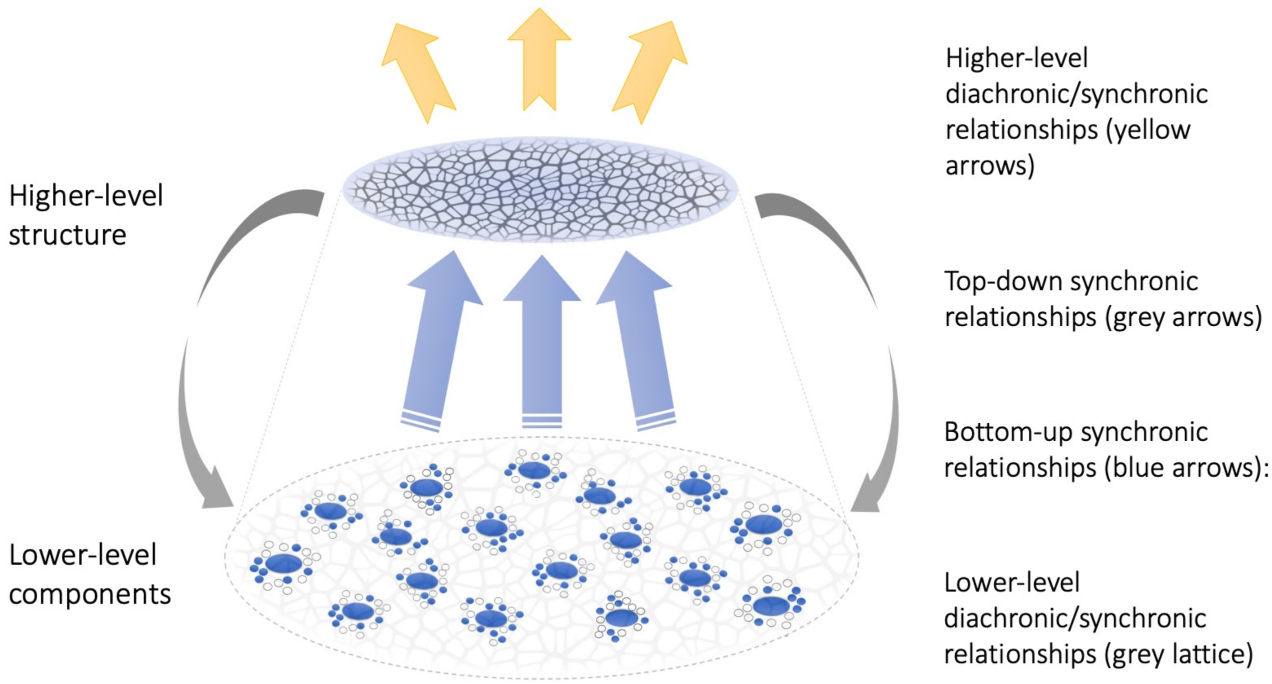

***Figure 3.1*. A visual representation of the structure of system exhibiting emergent properties.** The system is composed by a number of lower-level components inserted in a grey lattice representing their intrinsic powers and their basic low-level relationships, which realize or constitute (blue arrows) a higher-level organized structure. The structure, on the other hand, exerts a contraining influence (grey arrows) upon the components limiting their possible states (see Fig. 4.?? below). The structure, moreover, exhibits the emergence of novel forces able to influence the dynamics at the system level (yellow arrows).

conditioned way). In this framework it is worth noticing that despite the causal efficacy of the higher-level structure and its downward determinative influence upon the components, the properties and causal powers of the latter are not ontologically neutralized or replaced by the former, for the basic fundamental forces (the grey lattice) remain always the same, and consequently there is no conflict between the virtual exertion of basic causal powers and the formation of such constraining structures.

The properties of the parts that are wrapped in the system, therefore, are not compromised from a metaphysical point of view, but the effectiveness of the structures cannot be ignored at a systemic level (nor in general) because the constraints have decisive effects in filtering the possible states of the single components and the possible relationships between them in order to maintain the functionality of the whole or achieve systemic objectives. These constraints *make a difference* because, if they did not exist, the components would behave

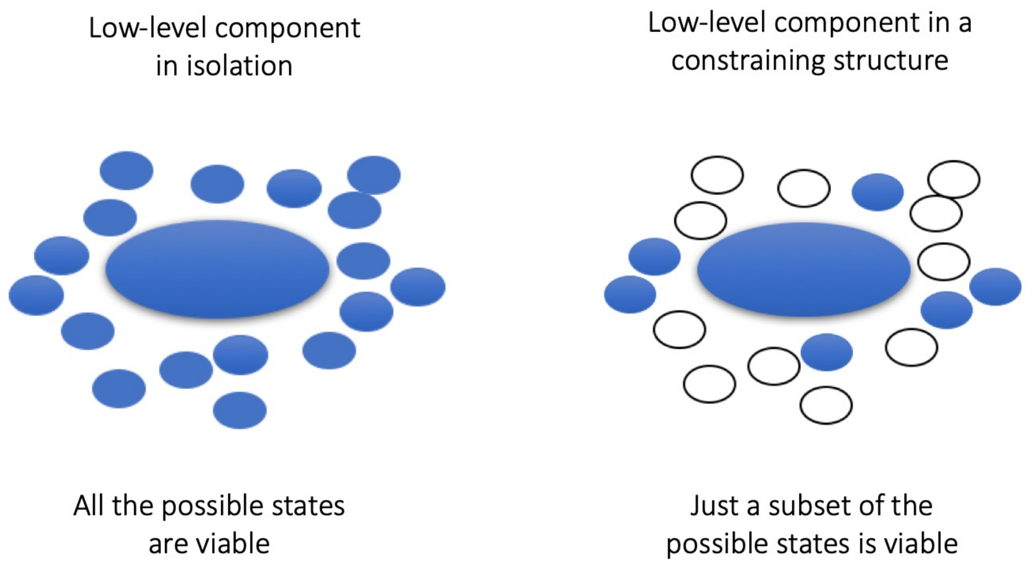

Figure 4.2. **The different viable possible states of a low-level component.** The difference between the viable possible state of a component in isolation (left) and in a constraining structure (right). In the first case the component is free, so all its possible states are viable (blue). In the second case, the component in constrained, so its viable state become a subset of its possible state: the subset composed by the states that are consistent with the functionality of the structure by which the component is constrained.

differently and reality would be very different as well: if there were no constraints the whole domain of biological organisms, for example, would be completely different in nature.

In this frame, moreover, the appearance of systemic constraints can represent the mark of a new level of organization:

> The formation of a new (relatively) macro constraint, however brought about, creates a new level proper in the system, since the constraint now filters out microscopic detail incompatible with it. The iron crystal lattice, e.g., filters out thermal fluctuations and many external perturbations, dissipating their energy as lattice vibrations. (Otherwise the constraint would not be stable against microscopic-originated perturbations and similar external disturbances.) The iron becomes a 2-level system, (1) that below the level of the lattice, the individual ions and electrons, obeying their dynamical interaction laws, and (2) that at the lattice level with its fermi conduction band where electrons stream through the lattice, the lattice collectively vibrates, and so on.[108]

[108] *Ivi*: 27.

This characterisation of levels as constraints-dependent is coherent with the piecemeal and local view described in §3.3.3. Levels can be detected just in particular structures and mechanisms, for the localisation of an entity into a specific level depends upon its role into the system to which it belongs. Levels should not be viewed as monolithic, context-insensitive layers of reality, therefore, but correspond to the particular differentiations in structure of the systems at issue. In those systems in which the parts are bounded by a global constrain, an at least two-level system can be described, and in this system the parts contribute with their own causal powers in just the ways admitted by the systemic constraints.

In this frame, therefore, emergent properties are often those systemic, constraining properties that influence their components without disrupting their own laws and properties, but just limiting them to a particular subset, which allows for the functionality of the whole. Given the restrictions on these delimited subsets of behaviours, the whole can achieve novel goals the unconstrained system could not, and this circumstance produces relevant differences in the nature and behaviour of the system, as well as in the environment in which it is found. Emergent phenomena, therefore, seem to present a twofold relational profile: on the one hand, they are involved in vertical synchronic relationships of constrain (or machresis) with their components, and, on the other hand, they are engaged into horizontal, high level causal or determinative relations.

To sum up, in addition to individual, active causal powers exerted by the basic physical entities composing reality, other determinative powers should be taken into consideration to understand and explain the development of natural systems. These other forms of causal determination are *diffused* because qualitatively realised by structures and not functionally realised by particular individuals, and they are able to regulate the causal powers exerted by these more basic entities through constraints. This role-*shaping* ability of emergent properties, which can be opposed to a role-*filling* nature which is typical of basic physical properties, requires a widening of the notion of causal efficacy. This is the reason why my suggestion is to integrate the account of causation based on individual power exertion with an account of causation including these further forms of determination, keeping in mind a wider meaning of the notion of causal efficacy. Being causally efficacious corresponds to the ability to bring about changes in reality, whereas the specific way in which that happens (exerting direct, intrinsic powers or not) should be a secondary question.

In this wider frame, becomes clear why emergence and reduction are not incompatible, being phenomena involving different kinds of properties and powers: on the one hand, basic, role-filling causal powers, and, on the other hand, constraining, role-shaping, determinative powers. These different kinds of power can and do coexist without being incompatible with each other, and they are both responsible of the complex structure of our world.

Conclusions

In this work I have tried to clarify what emergence is by highlighting the relationships between this concept and other concepts such as epistemological and ontological irreducibility, fundamentality, qualitative and causal novelty, and complexity. Emergence is often characterized as a state or process responsible for the appearance of phenomena that (i) cannot be ontologically reduced to their lower-level components, that (ii) cannot be deductively predicted with knowledge of these components, and that (iii) show new properties endowed with new types of causal or determinative efficaciousness. It should now be clear, however, that these three conditions cannot be taken into consideration without further investigation, given the different possible meanings that irreducibility and novelty possess.

The examination of these concepts and the understanding of their various implications might suggest that for each of the given interpretations a corresponding emergence model can be formulated. This is why it is possible to discuss so many types of emergence: epistemological, ontological, weak, strong, synchronic, diachronic and so on. This is because once the criteria have been determined, the definition will be in accordance with them. However, given the abundance of criteria (or rather the plurivocity of recognised criteria), there are many definitions that are only partially superimposable, and this multiplicity of descriptions has often been seen as an inherent weakness of the concept of emergence. This opinion, however, comes with the idea that this phenomenon can be defined once and for all through the identification of some specific characteristics that would arise at every instantiation of it.

What I have suggested in this work expresses a different opinion which should be clear by now. I do not think that emergence is a particular and defined phenomenon or process that occurs in different domains of reality in the same way. Emergent phenomena are heterogeneous, they have different characteristics according to the domain of reality in which they appear and according to the properties of the components of their emergence basis. The transformation of a fundamental particle into another type of particle, for example, is considered by Paul Humphreys a typical instance of ontological emergence. The author considers unstable leptons that transform into lighter leptons, neutrinos, quarks and antiquarks, and states that these processes "count as a type of emergence because they have three of the four characteristic features of emergence".[1] The new particles, in other words, (i) emerge from other types of entities, (ii) are autonomous and (iii) new. The fourth characteristic, holism, is not present, but this is not a big problem because for Humphreys none of the four characteristics is necessary.[2] The typical model of ontological emergence, for Humphreys, is therefore the transformational and diachronic model, a model in which elementary entities such as leptons are transformed into one another, leading to the disappearance of the non-emergent properties from which they emerged. Now, the relevance of the latter feature, i.e. the disappearance of lower-level properties as a result of the emergence process, depends however on the fact that the emergence examples analysed by Humphreys concern fundamental physical particles, which have a particular ontological nature. This characteristic cannot be extended to every case of emergence as its necessary and/or sufficient condition (and indeed Humphreys does not do so), since, rather obviously, not all entities in the world behave like subatomic particles. The case of ants, for example, cannot be seen in terms of transformational emergence because ants, unlike quantum particles, continue to exist, preserving their individual properties, even if organized in a colony.

The fact that the example given by Humphreys and the one described here do not seem to conform to the same set of conditions for emergence reflects a more general condition. The examples of emergent phenomena provided by the literature are so many and show such a heterogeneity that it is clear that it is impossible to define a few criteria sufficient to grasp the emergence phenomenon in a general and exhaustive way. Faced with this circumstance one can react in different ways, as I have anticipated in the introduction. On the one hand, one can adopt a *selective* strategy that determines a certain (tendentially small) number of criteria. This strategy risks becoming prescriptive, however, as depending on the criteria

[1] Humphreys 2016: 67.
[2] Ivi: 68.

identified some emergent phenomena will be included in the category and others will be excluded. A second strategy, motivated by the difficulties of the first one, could be the one I have defined as *liberal*, whereby, ideally, given that different types of emergence respond to different sets of conditions, one could assume all these sets as equally necessary. However, this strategy is too binding and faces the concrete risk of outlining an ontology of emergence that will prove to be deserted. As we have seen, in fact, emergent phenomena exhibit different characteristics depending on the ontological domain in which they occur and do not give rise to completely equal emergent phenomena at different levels of organisation. A strategy that requires emergent phenomena to exhibit all the characteristics identified in the literature (a literature ranging from physics to biology to the sciences of the mind) therefore appears clearly inadequate.

In the face of these difficulties and given the wealth of studies dedicated to emergence and the intrinsic complexity of the natural world, a change of perspective could be promising. Two complementary paths thus open up. On the one hand, that of directing the debate towards the formulation of a *specific disciplinary and subdisciplinary taxonomy*. In other words, it could be fruitful to clarify what are the most common characteristics of emergent phenomena in each of the fields where emergence is recognised, without mixing their characteristics in different ontological regions. There will therefore be specific standards for emergent phenomena that can be detected in the various subatomic, atomic, chemical, biological (and so on) domains, and what is distinctive in one area need not necessarily be distinctive in another - indeed, it would be rather unusual if this were the case. This road therefore consists in the formulation of several models able to capture the specificities of emergent phenomena in each particular ontological domain, a method that differs from the generalist one, aimed at identifying characteristics common to all cases of emergence.

This last operation is still possible, in my opinion - and this is the second path. We can identify a series of very general criteria for emergence in the awareness, however, that those criteria will not necessarily be exhaustive, that they can be integrated in the future and that, although useful to outline the phenomenon in general terms, they cannot be used in specific cases without examining their contextualization. The notion of emergence used in the debate on the nature of spacetime, for example, is different from that used in the analysis of complex adaptive systems and the latter, in turn, is different from the notion of emergence used by philosophers of the mind to talk about states of consciousness. In this framework, therefore, it seems promising to define emergence by resorting not to one or more sets of sufficient and necessary conditions, but to an open cluster of properties of which none should be understood as indispensable in every single case of emergence. But how should this cluster be characterised without falling into an excessive vagueness? And how can it be useful to use this conceptual tool to describe emergence effectively?

Now, the notion of cluster has been central to the debate on the meaning and reference of names, and has been exploited by scholars such as Ludwig Wittgenstein,[3] Peter F. Strawson[4] and John R. Searle.[5] These authors, together with others such as Frege and Russell, shared a descriptivist theory of proper nouns, according to which "the speaker associates a description to the name, which gives its meaning and determines its reference".[6] If we want to determine, for example, what the name "Socrates" refers to, we could say that it refers to the person who was "the master of Plato and Xenophon". The proper noun Socrates therefore refers to that individual who instantiates the property of being the master of Plato and Xenophon. Defining the meaning and/or reference of names[7] by means of a single description, however, may not be sufficient because there may be several entities that instantiate the property indicated by the chosen description (Plato and Xenophon may have had more than one master, for example). For this reason, in 1958, Searle proposed a revision of the theory, suggesting that names are not defined by a single description, but rather by a disjunction of descriptions. Associated with each name there would then be a family, a group or a *cluster* of properties, and the instantiation by an entity of a subset of these properties would make the entity in question the reference of the name.

A famous case that exemplifies this intuition is the analysis of the term 'game' carried out by Wittgenstein in his *Philosophical Investigations*.[8] Here, Wittgenstein notes that the term applies to cases that do not share the same set of properties, but rather resemble family members: “each has one or more features in common with some other member, but it cannot be said that there is a collection of traits that everyone shares”.[9] Terms such as 'game' refer to cases that present ‘family resemblance’ and in order to grasp the importance of these similarities Wittgenstein provides a significant methodological warning:

[3] Wittgenstein 2009.
[4] Strawson 2002.
[5] Searle 1958.
[6] Marino 2017: 8.
[7] The distinction "between using this theory as a theory of meaning and using it as a theory of reference" was underlined by Saul Kripke in *Naming and Necessity*. See Kripke 1980: 259 ff.
[8] Wittgenstein 2009: §§ 61-77.
[9] Paternoster 2001: 178.

> Don't say: "They must have something in common, or they would not be called "games" – but look and see whether there is anything common to all. For if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look![10]

This suggestion, in my opinion, is valuable. "Don't think but look" reflects the methodological position that I have expressed by favouring a metaphysically *descriptive* approach over a *prescriptive* approach to the problem of defining and understanding emergence. Similar to games, the notion of emergence is difficult to ultimately define given its intrinsic complexity, its dependence on the context and its many facets. As in the case of games, it is possible to impose conceptual limits that set necessary and sufficient conditions, but doing so risks making an arbitrary choice that includes some specific cases of emergence in the categorisation at the cost of excluding others. As Wittgenstein suggested with regard to games, instead, some definitions can be *open* because certain phenomena are too heterogeneous to be conclusively defined. Now, the concept of cluster fits perfectly into this conceptual framework. To grasp its characteristics, I will analyse two theories that refer to different phenomena. The first is the one formulated by Richard Boyd about natural kinds in biology, while the second is the prototype theory formulated by Eleanor Rosch to solve the problem of categorisation in psychology. Let's look at these theories and what suggestions they can provide to clarify the problem of defining emergence.

5.1 *The Homeostatic Property Cluster theory*

As we have anticipated, Richard Boyd applied the notion of clusters to the problem of natural kinds in biology and formulated the so-called Homeostatic Property Cluster (HPC)

[10] Wittgenstein 2009: 31. The passage goes on as follows: "Look for example at board-games with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost.-- Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. Look at the parts played by skill and luck; and at the difference between skill in chess and skill in tennis. Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared! sometimes similarities of detail. And we can go through the many, many other groups of games in the same way; can see how similarities crop up and disappear. And the result of this examination is: we see a complicated network of similarities overlapping and cries-crossing: sometimes overall similarities."

theory. The theory states that many relevant natural kinds, such as biological species, are defined by open clusters of properties and relationships rather than by sets of essential properties reflected by necessary and sufficient sets of conditions. The limitation of this second view – which is the traditional one[11] – lies in the fact that while the essential properties are immutable and universal, natural kinds such as biological species change over time and do not seem to group exactly identical individuals who instantiate all and only certain properties. One of the central problems of essentialism, which makes it inapplicable to biology,[12] is therefore its incompatibility with the theory of evolution and genetic variability. In the face of this and other difficulties,[13] Boyd then elaborated HPC, according to which for every natural genus there is a set of distinctive properties which are neither individually nor jointly necessary, but which nevertheless allow us to define (sometimes in a conditional form) the natural kind in question. *Homo sapiens*, for example, is defined by a set of properties such as 'being a mammal', 'being a plantigrade', 'being a biped', 'reproducing sexually', 'having forty-six chromosomes/thirty-two teeth (as adults)/four limbs/two kidneys/one uterus if female, etc. Yet, there are individuals of the *Homo sapiens* species who do not instantiate all the properties listed above – just think of those affected by chromosomal abnormalities (trisomy or monosomy) or by conditions for which some organs are missing (renal agenesis, dental agenesis or amelia, for example). Despite these anomalies, excluding these individuals from the *Homo sapiens* species would be unreasonable, but why? The answer to this question is twofold: on the one hand, there are more and less salient cluster properties, so not having an arm or a kidney does not seem to be such a significant anomaly as to compromise the person's belonging to the species (perhaps this depends on the fact that in the life of a *Homo sapiens* it may happen to lose a body part due to accidents or diseases, so this type of condition is seen as a more or less common possible condition). Things would be different if an individual of the *Homo sapiens* species were able to reproduce by parthenogenesis, for example; but this is only an unlikely thought experiment, because if it is true that species change and individuals can be born presenting genetic mutations, these are gradual mutations involving small parts of DNA, while a system like the sexual one cannot change suddenly from one generation to another. On the other hand, there is another reason why individuals with anomalies are still recognized as *Homo sapiens*: they do not instantiate all the properties of the cluster, but they

[11] See Borghini and Casetta 2012
[12] But the same could be said for chemistry, see Borghini and Casetta 2012.
[13] On this see Borghini and Casetta 2012.

do instantiate a considerable number of them. This last consideration raises the problem of the cluster content, and at this point one wonders how it can be established. In other words: how do we decide which properties belong to the cluster and which do not? How do we decide which are more or less salient? Is it a convention?

Boyd has a realist position about natural kind, so the author believes that the properties that make up the cluster are not selected arbitrarily or conventionally. They are real properties that co-occur permanently and regularly in organisms or natural systems by virtue of the presence of precise causal mechanisms and/or homeostatic structures. The properties that converge in a cluster therefore exist in this grouped form *already in the world* and not only in the minds of scientists and philosophers who seek to define natural kinds.[14] Nevertheless, when *defining* a cluster, there are heuristic considerations that influence the identification of the most salient properties and this interest depends on the theory that is assumed: there are therefore some epistemological factors that contribute to the definition of clusters (but not to their formation in the world). If the properties that form a cluster are therefore real properties already grouped in nature, the salience that is attributed to them in order to understand whether or not a certain entity belongs to it depends instead on the particular theoretical interest of the observer. This is why I consider the notion of cluster so promising to define emergence: on the one hand, the characteristics that distinguish emergent phenomena are co-occurring due to the relational structures and organizational forms that these phenomena exhibit: epistemological irreducibility due to non-linearity or qualitative novelty, ontological irreducibility due to causal novelty, robustness due to self-organization, autonomy due to robustness, and so on.

However, the ratio that defines which characteristics are more or less salient for the classification of each single case of emergence depends on the ontological domain in which it manifests itself: why and how can spacetime be called emergent? Why and how does quantum correlation produce emergent entities? Why and how can an ant colony manifest emergent behaviour? Why and how does the mind emerge from the nervous system? All these phenomena are different and it is starting from this differentiation that we need to formulate the disciplinary and subdisciplinary taxonomy we were talking about: a taxonomy that reflects the specificities of emergence and its manifestation in different ways in different ontological contexts. This taxonomy is disciplinary, but also sub-disciplinary because if it is true that there are considerable differences between the phenomena studied by physics

[14] Boyd 1991: 129.

and those studied by biology, it is also true that within the same disciplines there are as many differences: the emergence of space-time is not the emergence involved in quantum correlation, just as the emergence of pigmentation patterns similar to those of cellular automata on the shells of certain marine gastropods is not similar to the emergence of organised behaviour in social insect colonies. For every discipline, therefore, there will be one or more peculiar types of emergence that manifest certain distinctive traits, and all these traits will converge into the more general cluster made up of all the properties exhibited by emergent phenomena and that lead us to call apparently very different phenomena 'emergent'. At this point, however, it is evident that the internal ordering of the cluster must be dynamic, in the sense that there are no more or less salient properties in an absolute sense, but only with reference to a particular discipline or ontological domain. To better express this idea we can think about the prototype theory formulated by Eleanor Rosch, an American psychologist who in the 1970s investigated the processes of categorisation.

5.2 *The prototype theory*

Prototype theory contrasts with the classical view of categorisation in the cognitive sciences according to which the members of a category all share the same properties and therefore all have the same status within the whole. Rosch's theory suggests that the members of a category are not equal and that within each category there are asymmetries, called *prototype effects*. In other words, in each category, there are members who have a particular cognitive salience, so that if one asks a group of people to determine which are the best examples of sports, furniture, birds, fruit or vehicles, for example, one can reasonably anticipate that the answers will include football, chairs, sparrows, apples and cars much more often than archery, lamps, penguins, figs and tricycles.[15] The experiments carried out by Rosch and colleagues therefore indicate that within each category there are more and less representative members, which implies that each category has an internal radial structure[16] – like a Target – involving one or more prototypical specimens in the centre and specimens that are less and less similar to the prototype as you approach the periphery, up to borderline cases whose membership of the category is doubtful. The prototype theory, initially formulated by Rosch, but supplemented in later years by other

[15] These are the results of Rosch's experiments on her students. See Rosch 1973.
[16] On this see Paternoster 2001: 182 ff.

scholars and then applied to linguistics as well, seems to imply that inter-category boundaries are vague, since belonging to a category depends, for each member, on the family resemblances it presents with the other members and, above all, with the prototype member. This idea has sometimes suggested that it is not only our psychological categorisation that is graduated, but that a member's belonging to a category is itself graduated in the sense that, taking the examples given above, one could say that a tricycle is 'less a vehicle' than a car or that a penguin or hen is 'less a bird' than a robin.

However, this interpretation is incorrect and Rosch has intervened in more recent times to point out that the data from her experiments do not support it. Penguins, chickens and robins are full-fledged birds and the bird category is well defined (by zoology) so there are no animals that fit more and others that fit less. The latter is not the only problem of the "standard" prototype theory[17] which, as we have seen, is not only due to Rosch. The theory has had alternating luck and after some years of great popularity a series of criticisms (many from Rosch herself, as George Lakoff accurately reconstructs[18]) have weakened its premises and implications. However, the idea that the categories present a centre/periphery structure with more or less typical traits and that there is some form of (epistemological and/or ontological) vagueness at their borders is in my opinion entirely sharable and infinitely better than the idea that for each natural phenomenon there is a precise number of sufficient and necessary conditions capable of defining it exhaustively. This is perhaps possible for the simplest phenomena, such as electrons (an electron necessarily requires a charge, a mass and a spin), but it becomes less and less probable for more complex phenomena such as chemical compounds (is water really always and only H_2O?[19]) or, obviously, biological organisms.

Now, emergent phenomena are extremely heterogeneous and manifest themselves at different scales and at different levels of organisation and complexity, instantiating more or less specific properties. This heterogeneity means that the category of 'emergent phenomenon' can only be defined by an open cluster of properties whose internal structure is dynamic in the sense that all the properties it contains are in some sense (proto)typical, but depending on the specific ontological domain and on the kind of observer. The general cluster of emergence emerges in turn, we could say, from the overlapping of all the

[17] Alfredo Paternoster provides a six-point description of the standard version of the theory. See Paternoster 2001 and the authors cited by him in this regard, Kleiber 1990 e Violi 1997.
[18] See Lakoff 1987: 39 ff.
[19] See Weisberg 2006 and VandeWall 2007.

particular clusters that define the emergent phenomena typical of certain ontological regions. In the next paragraph I will take up the examples of emergent phenomena that I have outlined in this work to highlight their similarities and differences. The aim of this operation is to show through examples how different and at the same time similar emergent phenomena can be, how much their diversity depends on the ontological domain in which they appear and how much their similarity depends on the natural inclination of matter to structure itself in differentiated, complex and always new forms.

5.3 *(Proto)types of emergence*

Let us now take up the examples of emergent phenomena that we have described in the text, identifying the typical properties that justify their inclusion in the emergence category. As we will see, all these phenomena exhibit, at the right level of abstraction, some common characteristics, but also show significant differences, which suggests that adopting a classification which requires the fulfilment of the same conditions in each situation is inadequate to grasp the specificities of the various forms assumed by emergence in different ontological contexts. While it is true that there are similarities and that certain properties are shared in a stable way, it is also true that the various emergent phenomena instantiate those same properties in different ways and this suggests that it may be an appropriate move to take into account a cluster of properties with a variable internal structure. Let's now go over the characteristics of the emergence examples reported in this work.

In the domain of physics there are numerous examples of emergent phenomena. The first ones we have mentioned are quantum entanglement and particle decay, cases on which Paul Humphreys has focused for years. Humphreys, as we have seen, has formulated a model of *transformational* ontological emergence, identifying a series of typical properties that entangled systems and decayed (or 'transformed', according to his terminology) particles instantiate. These properties are (i) dependence (on an emergence base), (ii) autonomy and (iii) novelty, which in Humphreys' theoretical framework is always relative and never absolute. The emergence model capable of describing this type of phenomena is a diachronic one, which refers to processes and transformations that take place over time and that produce irreducible, new and autonomous phenomena. The emergence basis, following its *transformation* or *fusion*, ceases to exist and consequently the emergent phenomena are dependent on it, but only diachronically and not synchronically. This last circumstance also makes them authentically efficacious from a causal point of view: since there is no longer their emergence basis, the properties and the causal powers they instantiate are in no way inherited, but are proper to the emergent entities.

Another example of an emergent phenomenon in physics is spacetime, but although once again we are dealing with the subatomic domain, Humphreys' transformational emergence model does not seem adequate to describe its origin. As we have seen, according to authors such as Crowther and Wüthrich, spacetime can be said to be emergent because (i) it is derived from or dependent on a more fundamental basis, (ii) it is qualitatively new or heterogeneous, and, consequently it is (iii) autonomous (because it is robust and universal). The similarity with Humphreys' model is only apparent. It is true that all three authors recognise dependence, novelty and autonomy as typical traits of emergence, but on closer inspection there are significant differences between them. The most evident one concerns dependence: for Humphreys dependence is a simple sign of relationality, in the sense that emergence is relational and implies a relationship between an emergence base and a phenomenon emerging from it. However, this relation is diachronic and is configured as the transformation of the emergence base, which no longer exists once the emergent phenomenon appears. Crowther and Wüthrich, on the contrary, qualify dependence as a derivative or reductive relationship, so that within their framework emergent theories and phenomena are derivable and reducible to the most fundamental theories and structures (which continue to exist): a circumstance that is difficult to reconcile with Humphreys model, which also considers the failure of reductionism one of the necessary conditions for ontological emergence. Secondly, the transformational model is diachronic, while the emergence of spacetime cannot be defined in temporal terms because what emerges, together with space, is exactly time, understood as the time of relativity, but also as the phenomenological time of our everyday life.[20] To imagine a *before*, as well as a *during*, regarding the emergence of spacetime is clearly absurd, therefore this phenomenon cannot fall either into the category of diachronic emergence or into the category of synchronic emergence (unless we review the meaning of these categories, as suggested by Crowther and Sartenaer).[21]

The concept of emergence in the domain of quantum gravity and relativity, therefore, refers to a relationship of contemporary dependence and independence between fundamental and emergent structure, and this partial dependence, reminiscent of the one

[20] See Wuthrich in Gibs, Hendry, Lancaster 2019.
[21] See Sartenaer in Onnis 2019 and Crowther 2019.

outlined by Mark Bedau, is due to the appearance of qualitative novelty, which in this case corresponds to the spatio-temporality of emergent structures. The model formulated by Crowther and Wüthrich is therefore different from the transformational model elaborated by Humphreys – which seems to work for those phenomena involving subatomic particles – but also from the standard interpretation of emergence which, as the authors note, contrasts emergence with reduction.

Now, despite these differences, it seems appropriate to define spacetime as emergent by virtue of the sharing of some relevant cluster properties such as qualitative novelty, autonomy and robustness. The properties that are *not* shared are equally relevant, but the specificity of the ontological domain in question justifies these absences and anomalies. That same specificity, on the other hand, did not allow to place the emergent spacetime in that standard taxonomic grid that distinguishes synchronic emergence from diachronic one, as we have seen, nor in the one that distinguishes between epistemological and ontological emergence. It seems in fact that spacetime is emergent in both senses, since the issue involves both the relationship between the theories and the structures to which the theories refer. However, both Crowther and Wüthrich prefer to speak of weak or epistemological emergence because they observe that in philosophy strong or ontological emergence is assimilated to models that exclude the derivation between theories and the reducibility between entities. However, if the original sense of the distinction between epistemological and ontological emergence is to distinguish between a metaphysically innocent emergence, and a phenomenon that affects the structure of reality, it seems absurd to deny ontological relevance to the emergence of spacetime because of such categorical rigidities. It seems more sensible to admit that within the specific domain of subatomic physics, among the properties that constitute the cluster of emergence, qualitative novelty, the autonomy of the less fundamental phenomenon and its dependence on a more fundamental basis are sufficient to guarantee its emergence, where the impossibility of epistemological derivation and ontological reduction do not constitute necessary conditions. However, it should be clarified that the reducibility required by the emergentist theories of spacetime is not the universal one often addressed by philosophical reductionism, but the "partial" one, typical of science and now widely accepted among the philosophers of physics: a reduction that refers to approximation procedures, to the presence of limits or to effective field models and theories. This model of reduction is not, in my opinion, incompatible with emergence, as I have explained in the second chapter.[22]

We have seen, therefore, that even within the same discipline - physics - there are at least two emergence models which, despite having different characteristics, share some relevant cluster properties such as dependence, autonomy and novelty. I think it is clear, however, that the reason why there are different models of emergence does not depend on the inadequacy of our analysis, but on the specificity of the ontological material that the models have to configure: particles in Humphreys's diachronic transformational model, non-temporal structures in Crowther and Wüthrich's model. The two models are not the same, but not incomparable either, and it should be noted that although they are not superimposable as far as the negative criteria of irreducibility are concerned, they are quite superimposable as far as the positive criteria of novelty are concerned.

In addition to these two emergence models, we have seen others. The one described by Mark Bedau, called weak emergence, is a metaphysically innocent model of emergence that describes the dynamics of systems such as Cellular Automata (CA). These systems exhibit emergent patterns at the macro-level depending on the micro-level, in the sense that these superior patterns are generated by the non-linear dynamics of the micro-level components, but exhibit autonomy in being incompressible, i.e. in requiring a complete simulation to be reconstructed, and in being described by principles that do not take into account microscopic details. In other words: it is true that high-level patterns are exhaustively determined by low-level dynamics, but there is no way to predict what those dynamics will be based on their knowledge and the initial conditions of the system. The properties of weak emergence identified by Bedau are therefore, again, a certain type of dependence, which is generative and constitutive, and a certain type of autonomy which depends on the non-linearity of the CAs. This implies, on the one hand, the impossibility of predicting their dynamics without making a simulation and, on the other hand, the possibility of formulating high-level principles and descriptions for which microscopic details are irrelevant.

Also in this case of emergence there are some recurrent traits, but if we analyze in what terms they are outlined, we notice some inconsistencies and apparent incompatibilities: dependence, non-linearity and autonomy appear to be typical traits of emergence, but Bedau defines them in a way that enables the complete deductive predictability of the emergent properties (even if it is a predictability in principle that always requires simulation). Bedau

[22] In addition to the articles already mentioned, see also Buttefield 2011 and Crowther 2016.

himself wonders whether his emergence model is too weak to identify genuinely emergent phenomena: the absence of new causal powers and complete epistemological and ontological reducibility seem to be major obstacles. Yet, according to Bedau, the autonomy of higher level patterns is sufficiently marked to justify the attribution of an emergent character to these phenomena.

At this point I would like to add, however, that the lack of genuinely new powers may not entirely rule out high-level effectiveness because we have seen that, in the extended view of causality proposed here, causal-determinative effectiveness may occur even in the absence of intrinsic individual powers and forces. In order to understand if this is the case and if the high level patterns really have some kind of effectiveness (if they, in other words, produce ontological differences in the world) we can lower the dynamics of a CA into reality and analyse its effects and functions. Now, the classic examples of biological structures that seem to follow the rules of development of cellular automata are the shells of some marine molluscs that exhibit patterns extremely similar to those of cellular automata. The example that I would like to refer to, however, concerns reptiles and, in particular, a species of lizard, the ocellated lizard (*Timon lepidus*), which in recent years has been the subject of an accurate study[23] aimed at understanding the origin of its pigmentation patterns (Fig 5.1). The study determined that this lizard exhibits patterns derived directly from non-linear dynamics of cellular interaction at the mesoscopic level of its scales. However, these emergent patterns, created and generated by lower level dynamics (and derivable through simulation), have several biologically relevant functions. In males of this species, which

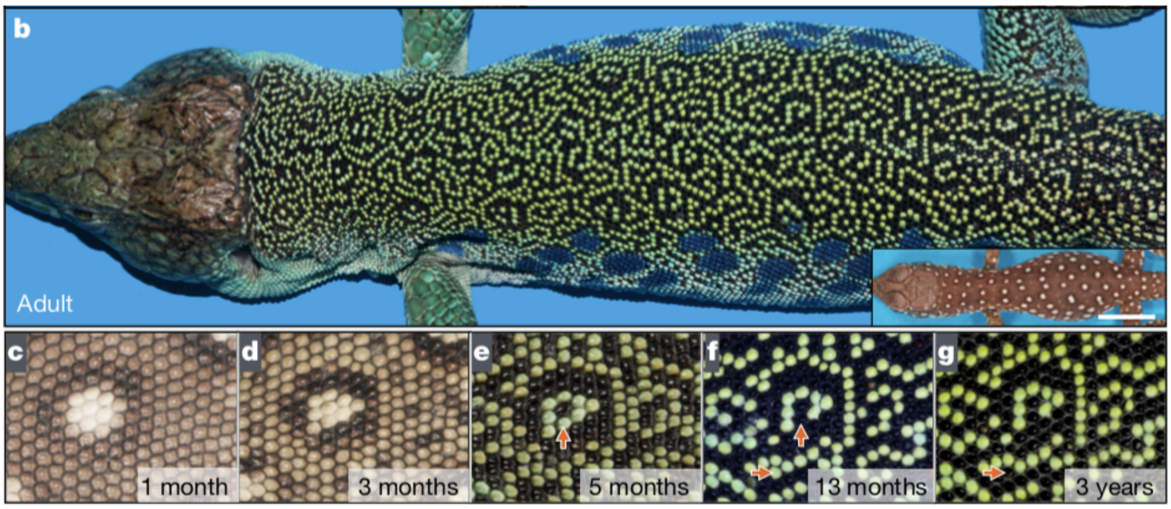

Figure 5.1 Colour pattern ontogeny in ocellated lizards. (Manukyan et al. 2017: 174).

[23] Manukyan et al. 2017.

presents sexual dimorphism, the patterns represent, for example, a secondary sexual character that has a decisive influence in courtship, reproduction and, more generally, in the evolution of the species. The first question to ask now is whether patterns can be given a causal or determinative effect. My answer is yes: patterns have a real function and relevant effects at the macro level. The second question is whether this determinative capacity derives from particular powers intrinsic to the pattern. In this second case the answer is partially negative: the form of the pattern intrinsically contains information, but the determinative efficacy that the pattern exerts thanks to it is extrinsic since it depends on the environment, that is, on the ability of other individuals to process the codified information.

The emergent patterns that Bedau recognises in complex systems therefore enter the category of emergent phenomena by virtue of a series of properties that are partially different from those of the models analysed so far: there is dependence, there is non-linearity and therefore autonomy, but there is also derivability (but only in principle); it can also be argued that there is some form of high-level determinative effectiveness due to the qualitative novelty that the patterns incorporate. This example of emergence is interesting because the particular cluster of properties that characterize the specific phenomena we are talking about seems to include at the same time a certain type of reducibility and a different type of irreducibility. On the one hand, the emergent patterns are reducible to the microdynamics that generate them because they are completely determined by it, and the microdynamics, in turn, can be deduced from the knowledge of iys development rules and initial conditions. On the other hand, the qualitative novelty that patterns manifest (their form and the information encapsulated in it) remains ontologically distinct from the processes that produced it and proves capable of high-level determination - a determination that it acquires not only by virtue of its individual characteristics, but also (and above all) by virtue of its environment, the relational network in which it appears and its evolutionary history. If the patterns of the ocellated lizard transmit a message, in short, they do so because they have been selected during evolution, just as the same ability to "read" messages and information in general has been selected by evolution. The lesson we can learn from these considerations is that in order to define the determinative effectiveness of an entity it is not sufficient to analyse its individual powers, since effectiveness could depend on a complex network of relationships that can only be grasped by broadening the perspective and considering the context.

Finally, the last example I would like to come back to is the one I spent the most words on: ant colonies or, more generally, complex adaptive systems. These systems are made up of a large number of (often simple) elements interconnected by sophisticated relational structures that allow the system to reach a (dynamic) state of self-organization capable of influencing, on the one hand, the individual states of the components and, on the other hand, the phenomena that are at its own level of organization. In the case of ants, the components are the ants, their chemical and mechanical exchanges create the complex relational network, and the structure that emerges has the ability to interact with other ant colonies, with other species and with the environment. It also has the ability to influence the behaviour of the individuals that compose it through constraints that limit it to the subset of possibilities useful for the pursuit and maintenance of the functions and objectives of the colony. In this case, the systemic properties of the ant colony can be described as emergent phenomena by virtue of the partial dependence of the ant colony on individual ants, the autonomy and robustness of its behaviour (derived also in this case from the non-linearity of the underlying relationships), as well as its ability to exert causal-determinative effectiveness on its own components and on other entities at a macro-level.

Therefore, once again, we have dependence, autonomy and determinative effectiveness, but also a certain form of epistemological reducibility or predictability that is not only not incompatible with emergence, but is a direct effect of it. Let me explain. Ants living in a colony have a much more predictable behaviour than those that do not belong to one because the emergent systemic structure, in binding their behaviour, constrains them to precise systemic objectives and therefore "orders" their behaviour, which becomes more comprehensible to an observer. The predictability that characterises many complex adaptive systems, however, has a different root from that of CAs. In the case of the latter systems, predictability depends on the fact that each step in the evolution of the system is vertically and synchronously determined by the underlying microdynamics, so the possibility of predicting its evolution depends on the exact knowledge of this low-level microdynamics. In the case of complex systems, predictability depends instead on understanding high-level systemic objectives and not on local microdynamics because the system is *finalized*. In both cases high level dynamics are realised by low level ones, but if in the case of cellular automata only microdynamics is the driving force, in the case of complex systems microdynamics are forced and constrained by macrodynamics in a circular relationship of mutual determination.

5.4 *Emergent phenomena*

From this review we can conclude that being an emergent phenomenon corresponds to instantiating some of the properties of a cluster that is heterogeneous, open and always integrable. The properties of the cluster include properties that can be defined as negative and properties that can be considered positive, but in my opinion they weigh differently from an epistemic point of view. As we have said, irreducibility is a negative criterion: an entity is emergent if its properties cannot be deduced epistemologically or broken down ontologically into smaller, simpler and more fundamental components. Negative criteria, however, do not tend to answer the question "what is x?", but rather the question "what (or how) x is not?", so even if they are of great importance in delineating the nature of a phenomenon, distinguishing it from phenomena of a different type, they are insufficient because they do not provide direct indications of their positive characteristics. The negative criteria, in short, are useful to *identify* an entity, but they are inadequate to *define* it.

It should also be noted that many of these criteria have a relevance that is sometimes exaggerated. On the other hand, if it is difficult to define emergence positively, defining it by negation is not much easier because if it is not clear what traits are instantiated by emergent phenomena, it cannot be clear what traits never are. In fact, the notion of irreducibility, once analysed and specified, has proven to be not very effective in the exhaustive identification of emergent phenomena. If we take the functional one as a model of reduction, for example, all the properties that are not functional will be irreducible. If one assumes a model of irreducibility that is based on the notion of realization (and bases this notion on the performance of causal roles), all non-causal properties in the strict sense will be irreducible. One can certainly admit that all non-functional and non-causal properties are emergent, but this statement should be further substantiated, I believe. If one then assumes an epistemological model of Nagelian reducibility, the vast majority of the theories will be irreducible. If the model is revised, some theories will become reducible, but at the cost of admitting a certain degree of vagueness. Also in this case one could admit the presence of many cases of epistemological emergence (a hypothesis that I do not consider unreasonable), but even in these cases some more consideration seems to be needed. Finally, if we admit a milder model of reduction, which is the one that scientists actually exploit (the one that considers the notion of limit and field effectiveness), emergence does not seem to be affected at all because the model has no universal metaphysical implications, but on the contrary it seems to imply a certain intrinsic ontological differentiation detectable on different scales and at different levels of organization.

As anticipated in the second chapter, emergent phenomena therefore instantiate reducible properties and irreducible properties in the sense that they instantiate both properties that can be reduced (the functional and sometimes causal ones) and properties that are genuinely new, such as causal-determinative or qualitative ones. The pigmentation pattern structures of lizards, for example, are fully reducible to the dynamics of cellular interaction at the epithelial level, but their function and the information they convey is ontologically distinct from the dynamic ones - it is qualitatively different and therefore new. The qualitative heterogeneity of information therefore lies in its form and its semantic value, whose structure can be completely determined by physical dynamics, but whose identity is irreducibly different and new. The same discourse can be applied, I believe, to the question of consciousness, which I chose not to address within this dissertation for lack of space, but which can certainly be mentioned, albeit briefly, in these conclusions.

It is well known that the most cryptic aspect of the mind has to do with phenomenal consciousness, i.e. the qualitative aspect of mental states. A classic example is pain:

> […] there is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fiber firing is *why pain should feel the way it does*! For there appears to be nothing about C-fiber firing which makes it naturally “fit” the phenomenal properties of pain, any more than it would fit some other set of phenomenal properties. The identification of the qualitative side of pain with C-fiber firing (or some property of C-fiber firing) leaves the connection between it and what we identify it with completely mysterious. One might say, it makes the way pain feels into merely brute fact.[24]

The qualitative aspect of pain therefore seems to be a "brute fact", i.e. an emergent property. Although this hypothesis has triggered a broad and articulated debate, especially on the presumed causal autonomy of the mind, it seems reasonable to think that phenomenal states of consciousness are after all the clearest example of qualitative novelty and emergent heterogeneity. It is no coincidence that the mind was one of the emergent phenomena *par excellence* according to the British emergentists, who – remember – did not admit metaphysical dualism. The emergence base of consciousness, or its components, is therefore entirely physical and probably corresponds to the central nervous system. If we trust Gillett's interpretation, we can even say that the neurobiological basis *realizes* states of

[24] Levine, 1983, p. 357.

consciousness, which would not be mysterious or scientifically incomprehensible because they can somehow be reduced to a less complex level of organization. However, the qualitative states or *qualia* resist any reduction because they are ontologically heterogeneous with respect to their implementation/emergence basis. Qualitative novelty therefore seems compatible with forms of functional or causal reduction, determinism and epistemological reductionism.

Now, returning to the question of negative criteria, although they prove to be only partially efficacious, they reflect real traits that emergent phenomena undeniably present, so they rightly belong to the cluster that defines emergence as well as the positive criteria. As for the latter, they correspond, essentially, to the various ways in which an entity can be new, and this is a positive criterion because it attributes substantial characteristics to emergent phenomena. We have seen that novelty can be declined in different forms - non-linearity, fundamentality, qualitative and causal or determinative novelty. All these forms are part of the cluster that defines emergence together with irreducibility and many other more particular properties such as robustness, incompressibility, universality and so on.

Summing up, to describe emergent phenomena it is necessary to use a heterogeneous set of properties: a cluster that is not defined once and for all because, on the one hand, emergence occurs in different ways in different contexts and because, on the other hand, our knowledge about it is constantly increasing. Moreover, emergence is one of the ways in which matter is organised and this organisation is itself subject to evolution, so it cannot be excluded that new forms of organisation will appear in the future, just as new forms have appeared so far in the course of evolution. It is therefore advisable in many ways to leave the emergence cluster open and give it a flexible internal organisation, changing according to the ontological domain in which the emergent phenomena occur. This is perhaps the only way to recognize the heterogeneity of emergence: a heterogeneity that reflects the complexity of reality.

Acknowledgement

I would like to express my gratitude and appreciation to all those who have helped me, supported me, and bore with me during these years, allowing me to conceive, reconceive, reconceive again, and finally write – after some further reconceiving – this thesis.

First of all, I wish to thank the FINO Consortium and the University of Genoa, which funded me during these years, making possible my research and my visiting stays in Ireland and in Portugal, as well as the many travels I made to present my research and attend conferences, workshops, and summer schools. Thanks to the board of teachers, to the steering committee, and, obviously, to the administrative offices.

Then, I wish to thank those who read my work and followed me throughout the process of writing. Thanks to my supervisor, Alfredo Paternoster, for his dedicated support, his kind encouragement, and the wise advice he gave me throughout this project. Thanks for reading my work with so much attention and for guiding me towards more clear and neat formulations of my thoughts. Thanks to Joel Walmsley for his enthusiasm for the project, for the constant guidance, and for the time he devoted to me. Thanks for the genuine interest and for sharing with me so many theoretical and historical insights. Our discussions improved my work, making it richer and more complete. Eventually, thanks for the warm hospitality during the months I spent at the University College in Cork. Similarly, thanks to David Yates not only for the valuable feedback he gave me and for the useful discussions we had about why and how being emergentists today, but also for having invited me in Lisbon and having introduced me to his research group and to his project "Emergence in the Natural Sciences", funded by the Fundação para a Ciência e a Tecnologia.

Moreover, I wish to sincerely thank many of the authors I mentioned in this work, and in particular Jessica Wilson and Carl Gillett, for having discussed with me about their accounts of emergence, for having shared unpublished or published articles and books with me, and for having been curious and interested in my project.

Eventually, I am grateful to all those who spent their priceless time – hopefully not against their will – to discuss with me about emergent phenomena. During the last years, this happened so many times in person, by phone, on Skype, through endless voice messages or chats, during conferences, seminars, lessons, travels, coffee breaks, lunch breaks, dinners, and so on and so forth. Thanks to Tiziana Andina, Alessio Bucci, Francesco Camboni, Elena Casetta, Robert Clowes, Giovanni Leghissa, Marta Conti Lorenzo, Joao Cordovil, Karen Crowther, Maurizio Ferraris, Joaquim Giannotti, Robin Hendry, Vincenzo Idone Cassone, Jimmy Hernandez Marcelo, Claudio Marciano, Gloria Sansò, Gil Santos, Enrico Terrone, Davide Vecchi, Marco Viola, Niki Young, and all the students I had in the past three years (especially the reductionist ones) who challenged on a regular basis everything I said.

I am deeply grateful to you all. Thank you.

References

Adams, R. M. (1979). Primitive thisness and primitive identity. *The Journal of Philosophy*, 76(1), 5-26.

Aguirre, A., Foster, B., & Merali, Z. (Eds.). (2019). *What is Fundamental?*. Berlin: Springer.

Albert, R., Jeong, H., & Barabási, A. (1999). Diameter of the World-Wide Web. *Nature*, 401(6749), 130–131.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P., (2002) *Molecular biology of the cell* (4th ed.) New York: Garland Science.

Alexander, S. (1920). *Space, Time, and Deity*. London: Macmillan.

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

Anderson, P.W. (2000). Sources of quantum protection in high-Tc superconductivity. *Science*, 288(5465), 480-482.

Assad, A., & Packard, N. H. (1992). Emergence. In M. A. Bedau & P. Humphreys (Eds.). (2008). *Emergence: Contemporary readings in philosophy and science* (pp. 231–234). Cambridge: MIT Press; page citations refer to this printing. Originally appeared as Sect. 2 of "Emergent colonization in an artificial ecology". In F. Varela & P. Bourgine (Eds.), *Towards a practice of autonomous systems: Proceedings of the First European Conference on Artificial Life* (pp. 143–152). Cambridge, MA: MIT Press.

Atlan, H., & Cohen, I., (1998). Immune information, self-organization and meaning. *International Immunology*, 10(6), 711-717.

Bain, J. (2013). The emergence of spacetime in condensed matter approaches to quantum gravity. *Studies in the History and Philosophy of Modern Physics*, 44:338–345, 2013.

Baysan, U. (2015). Realization relations in metaphysics. *Minds and Machines*, 25(3), 247-260.

Barnes, E. (2012). Emergence and Fundamentality. *Mind*, 121:873–901.

Barnes, E. (1994). Explaining brute facts. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1994, No. 1, 61-68.

Barnes, J., ed. *The Complete Works of Aristotle*, Volumes I and II, Princeton: Princeton University Press, 1984.

Bateson, G. (1972). *Steps to an ecology of mind*. Chicago: University of Chicago Press.

Batterman, R.W. (2001). *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence*. Oxford: Oxford University Press.

Batterman, R.W. (2011a). Emergence, singularities, and symmetry breaking. *Foundations of Physics*, 41(6), 1031-1050.

Batterman, R. W. (2011b). The tyranny of scales. In R. Batterman, (Ed.). (2013). *The Oxford handbook of philosophy of physics*. Oxford University Press.

Batty, M. (2012). Building a science of cities. *Cities*, 29, S9-S16.

Bedau, M.A. (1997). Weak Emergence. *Philosophical Perspectives*, 11, 375–399.

Bedau, M. (2002). Downward causation and the autonomy of weak emergence. *Principia: an international journal of epistemology*, 6(1), 5-50.

Bedau, M. A. (2011). Weak emergence and computer simulation. *Models, simulations, and representations*, 91-114.

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

Bennett, K. (2017). *Making Things Up*. Oxford: Oxford University Press.

Berto, F. and Tagliabue, J., "Cellular Automata", The Stanford Encyclopedia of Philosophy (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/cellular-automata/>.

Bich, L., & Mossio, M., (2011). On the Role of Constraints in the Emergence of Biological Organization. *Logic and Philosophy of Science*, 9(1), 381-388.

Bickle, J. (2008). Real Reduction in Real Neuroscience: Metascience, Not Philosophy of Science (and Certainly Not Metaphysics!). In Hohwy, J., & Kallestrup, J. (Eds.). *Being reduced: New essays on reduction, explanation, and causation* (pp. 34–51). Oxford University Press.

Bird, A. (2010). *Nature's metaphysics: Laws and properties*. Oxford: Oxford University Press.

Bishop, R.C. (2003). On separating predictability and determinism. *Erkenntnis*, 58(2), 169-188.

Bishop, R.C. (2010), Metaphysical and Epistemological Issues in Complex Systems. In C. Hooker (ed.) *Philosophy of Complex Systems* (119–150), Vol. 10, *Handbook of the Philosophy of Science*. Amsterdam: North Holland.

Blachowicz, J. (2013). The constraint interpretation of physical emergence. *Journal for general philosophy of science*, 44(1), 21-40.

Bliss, Ricki and Trogdon, Kelly, "Metaphysical Grounding", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/grounding/>.

Blitz, D. (2013). *Emergent evolution: qualitative novelty and the levels of reality*. Berlin: Springer.

Block, N. (1980). Troubles with functionalism. Readings in philosophy of psychology, 1, 268-305.

Block, N. J. (1982). Functionalism. In Cohen, J. J., Los, J., Pfeiffer, H., & Podewski, K. P. (Eds.). (2011). *Logic, Methodology and Philosophy of Science VI*. Elsevier.

Bokulich, Alisa, "Bohr's Correspondence Principle", The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2014/entries/bohr-correspondence/>.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: from natural to artificial systems*. New York: Oxford University Press.

Borghini, A., & Casetta, E. (2012). 4.2. Quel che resta dei generi naturali. *Rivista di estetica*, (49), 247-271.

Brandman, O., & Meyer, T. (2008). Feedback Loops Shape Cellular Signals in Space and Time. *Science*, 322(5900), 390–395.

Bretagnolle, A., Pumain, D., & Vacchiani-Marcuzzo, C. (2009). The organisation of urban systems, in D. Lane, D. Pumain, S. Van der Leeuw, G. West, eds., *Complexity perspective in innovation and social change*, Berlin: Springer,197–220.

Brockman, D.J. (1996). *Third culture: Beyond the scientific revolution.* New York: Simon and Schuster.

Butterfield, J. (2011). Emergence, reduction and supervenience: A varied landscape. *Foundations of Physics*, 41:920–959, 2011.

Menon, T., & Callender, C. (2011). Turn and face the strange... ch-ch-changes: Philosophical questions raised by phase transitions. In R. Batterman, (Ed.). (2013*). The Oxford handbook of philosophy of physics*. Oxford University Press.

Camazine, S, Deneubourg, J-L, Franks, N, Sneyd, J, Theraulaz, G, & Bonabeau, E (2001). *Self-Organization in Biological Systems,* Princeton University Press, Princeton, NJ.

Cao, T.Y. (Ed.). (2004). *Conceptual foundations of quantum field theory*. Cambridge: Cambridge University Press.

Cartwright, N. (1994). Fundamentalism vs. the Patchwork of Laws. In *Proceedings of the Aristotelian Society* (Vol. 94, pp. 279-292). Aristotelian Society, Wiley.

Clark, A. (1996). Happy couplings: Emergence and explanatory interlock. In *The Philosophy of Artificial Life*, edited by M. Boden, 262–81. Oxford University Press, Oxford.

Clark, A. (2001). *Being there. Putting brain, body, and world together again*. Cambridge, Mass.: MIT Press.

Chalmers, D. J. (2006). Strong and weak emergence. In Clayton, P., & Davies, P. (2006). *The re-emergence of emergence* (pp. 244-256). NewYork: Oxford University Press.

Chaitin, G. J. (1979). "Toward a mathematical definition of 'life'", in *The Maximum Entropy Principle*, R. D. Levine and M. Tribus, eds., MIT Press, Cambridge, Massachusetts.

Chibbaro, S., Rondoni, L., & Vulpiani, A. (2014). *Reductionism, emergence and levels of reality*. Berlin: Springer.

Churchland, P. M., & Churchland, P. S. (1992). Intertheoretic reduction: A neuroscientist's field guide. In Churchland, P. S. (Ed.). *Neurophilosophy and Alzheimer's disease.* (pp. 18-29). Berlin: Springer.

Cohen, I.R. (2006). Informational Landscapes in *Art, Science, and Evolution*, *Bulletin of Mathematical Biology*, 68: 1213-1229. https://doi.org/10.1007/s11538-006-9118-4

Cook, M. (2004). Universality in elementary cellular automata. *Complex Systems*, 15(1):1–40.

Corradini, A., & O'Connor, T. (Eds.). (2010). Abingdon: Routledge.

Correia, F., & Schnieder, B. (Eds.). (2012). *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge: Cambridge University Press.

Crane, T. (2001). The significance of emergence. In C. Gillett & B. Loewer (Eds.), *Physicalism and its discontents* (pp. 207–224). Cambridge: Cambridge University Press.

Crook, S., & Gillett, C. (2001). Why physics alone cannot define the 'physical': Materialism, metaphysics, and the formulation of physicalism. *Canadian Journal of Philosophy*, 31(3), 333-359.

Crowther, K. (2013). Emergent spacetime according to effective field theory: From top-down and bottom-up. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3), 321–328.

Crowther, K. (2016). Effective spacetime: *Understanding emergence in effective field theory and quantum gravity*. Berlin: Springer.

Crowther, K. (2019). When Do We Stop Digging? Conditions on a Fundamental Theory of Physics. In Aguirre, A., Foster, B., & Merali, Z. (2019). *What is fundamental?* (123-133). Berlin: Springer.

Crowther, K., & Linnemann, N. (2017). Renormalizability, fundamentality, and a final theory: The role of UV-completion in the search for quantum gravity. *The British Journal for the Philosophy of Science*, 70(2), 377-406.

Cucker, F., & Smale, S. (2007). Emergent behavior in flocks. *IEEE Transactions on automatic control*, 52(5), 852-862.

Cunningham, B. (2001). The Reemergence of Emergence. *Philosophy of Science*, 68: S62–S75.

Dennett, D. C. (1991). Real Patterns. *The Journal of Philosophy*, 88(1), 27–51.

Daniel, D. (1995). *Darwin's dangerous idea*. New York: Simon and Schuster.

Dyke, C. (1988*). The Evolutionary Dynamics of Complex Systems: A Study in Biological Complexity*. New York: Oxford University Press.

Ekeland, I. (2002). *Le chaos*. Paris: Le Pommier.

Emden, M. H. van (1971). *An analysis of complexity*. Mathematisch Centrum, Amsterdam.

Érdi, P. (2008). *Complexity explained*. Berlin: Springer.

Detrain, C., & Deneubourg, J. (2006). Self-organized structures in a superorganism: do ants "behave" like molecules? *Physics of Life Reviews*, 3(3), 162–187.

Ferrari, M. (2015). *L'evoluzione è ovunque: vedere il mondo con gli occhi di Darwin*. Milano: Codice.

Ferraris, M. (2009). *Documentalità. Perché è necessario lasciar tracce*, Roma-Bari, Laterza; Engl transl. by R. Davies, *Documentality: Why It Is Necessary to Leave Traces*, New York: Fordham University Press, 2012.

Feyerabend, P. (1981). Problems of Empiricism: Philosophical Papers. Cambridge: Cambridge University Press.

Fine, K. (2012). Guide to Ground. In F. Correia & B. Schnieder (eds.), *Metaphysical Grounding*. Cambridge University Press. pp. 37–80.

Flake, G., Lawrence, S., Giles, C., & Coetzee, F. (2002). Self-organization and identification of Web communities. *Computer*, 35(3), 66-70.

Floridi, L. (2010). *Information – A Very Short Introduction*. Oxford: Oxford University Press.

Floridi, L. (2014). *The philosophy of information*. Oxford: Oxford University Press.

Floridi, Luciano, "Semantic Conceptions of Information", The Stanford Encyclopedia of Philosophy (Spring 2017 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/spr2017/entries/information-semantic/

French, Steven, "Identity and Individuality in Quantum Theory", The Stanford Encyclopedia of Philosophy (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2015/entries/qt-idind/>.

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

Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford: Oxford University Press.

Janeway, C., Travers, P., Walport, M., & al (2001). Immunobiology: the Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001.

Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, 223(4):120–123

Gibb, S., Hendry, R.F., & Lancaster, T. (Eds.) (2019) *The Routledge Handbook of Emergence.* Abingdon: Routledge.

Gillett, C. (2006). Samuel Alexander's emergentism: Or, higher causation for physicalists. *Synthese*, 153(2), 261–296.

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

Glasner, E., & Weiss, B. (1993). Sensitive dependence on initial conditions. *Nonlinearity*, 6(6), 1067–1075.

Gordon, D. (2010). *Ant encounters: interaction networks and colony behavior*. Princeton, NJ: Princeton University Press.

Gordon, D. M. (2016). The Evolution of the Algorithms for Collective Behavior. *Cell Systems*, 3(6), 514–520.

Grassé, P. P. (1959). La reconstruction du nid et les coordinations interindividuelles chez Bellicositermes natalensis et Cubitermes sp. La theorie de la stigmergie: Essai d'interpre´tation du comportement des termites constructeurs. *Insectes Sociaux*, 6(1), 41–83.

Guay, A., & Sartenaer, O. (2016). A new look at emergence. Or when after is different. *European Journal for Philosophy of Science*, 6(2), 297-322.

Hare, R. M. (1952). *The Language of Morals*. Oxford: Clarendon Press.

Hawley, K. (2006). Weak discernibility. *Analysis*, 66(4), 300-303.

Heath, T. L. (Ed.). (1956). *The thirteen books of Euclid's Elements*. North Chelmsford: Courier Corporation.

Hempel, C. G. (1950). Problems and changes in the empiricist criterion of meaning. *Revue internationale de philosophie* 4(11), 41-63.

Hempel, C. (1969). Reduction: Ontological and Linguistic Facets. In P. Suppes, S. Morgenbesser & M. White (Eds.), *Philosophy, Science, and Method: Essays In Honor of Ernest Nagel* (179-199). New Yorm: St. Martin's Press.

Hendry, R. (2006). Is There Downward Causation in Chemistry? In D. Baird, E.R. Scerri, & L.C. McIntyre, (Eds.). *Philosophy of chemistry: Synthesis of a new discipline*. Dordrecht: Springer.

Hettema, H. (2017). *The union of chemistry and physics: linkages, reduction, theory nets and ontology*. Berlin: Springer.

Hitchcock, C. (2018). Probabilistic Causation, *The Stanford Encyclopedia of Philosophy*, E.N. Zalta (Ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/causation-probabilistic/>

Hobson, A. (2013). There are no particles, there are only fields. *American Journal of Physics*, 81(3), 211-223.

Holland, J. H. (2014). *Complexity: a very short introduction*. Oxford: Oxford University Press.

Hollik, W. (2010). Quantum field theory and the Standard Model. arXiv preprint arXiv:1012.3883.

Hooker, C. (2011). *Handbook of the Philosophy of Science, Philosophy of Complex Systems*. Oxford-Amsterdam: Elsevier

Hooker, C. (2013). On the import of constraints in complex dynamical systems. *Foundations of Science*, 18(4), 757-780.

Horst, S. (2007). *Beyond reduction: Philosophy of mind and post-reductionist philosophy of science*. Oxford: Oxford University Press.

Hu, B. (2009). Emergent/quantum gravity: Macro/micro structures of spacetime. Journal of Physics: Conference Series, 174, 012015.

Hubbell, J. H. (2006). Electron–positron pair production by photons: A historical overview. *Radiation Physics and Chemistry*, 75(6), 614-623.

Humphreys, P. (1997a). Emergence, Not Supervenience. *Philosophy of Science*, 64, S337-S345. Retrieved from http://www.jstor.org/stable/188415

Humphreys, P. (1997b). How properties emerge. *Philosophy of science*, 64(1), 1-17.

Humphreys, P. (2016a). Emergence. In Humphreys P. (ed.) *Oxford Handbook of the Philosophy of Science* (pp. 759-778). New York: Oxford University Press, 2016.

Humphreys, P. (2016b). *Emergence. A Philosophical Account*. New York: Oxford University Press.

Huxley, T. H. (1864). Criticisms on The origin of species. *Natural History Review New Series 4*: 25-36.

Hüttemann, A. (2004). *What's wrong with microphysicalism?.* London: Routledge.

Hüttemann, A. (2005). Explanation, emergence, and quantum entanglement. *Philosophy of science*, 72(1), 114-127.

Johnson, S. (2002). *Emergence: The connected lives of ants, brains, cities, and software*. New York: Simon and Schuster.

Kanwal, M., Grochow, J., & Ay, N. (2017). Comparing Information-Theoretic Measures of Complexity in Boltzmann Machines. *Entropy*, 19(7), 310–326.

Kidd, B. A., Peters, L. A., Schadt, E. E., & Dudley, J. T. (2014). Unifying immunology with informatics and multiscale biology. *Nature Immunology*, 15(2), 118–127. http://doi.org/10.1038/ni.2787

Kim, J. (1984). Epiphenomenal and Supervenient Causation'. *Midwest Studies in Philosophy IX: Causation and Causal Theories*, 257–270.

Kim, J. (1993). *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

Kim, J. (1999). Making sense of emergence. *Philosophical studies*, 95(1-2), 3–36.

Kim, J. (2007). *Physicalism, or something near enough*. Princeteon: Princeton University Press.

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

Kirschner D. (2007) The Multi-scale Immune Response to Pathogens: M. tuberculosis as an Example. In: Flower D., Timmis J. (eds) *In Silico Immunology*, Berlin: Springer.

Kirwan, C. (1993). *Metaphysics: Books Gamma, Delta, and Epsilon*. Oxford: Clarendon Press.

Kitano, H. (2004). Biological robustness. *Nature Reviews Genetics*, 5(11), 826.

Kleiber, G. (1990). *La sémantique du prototype: catégories et sens lexical*. Presses Universitaires de France-PUF.

Kment, B. (2010). Causation: Determination and Difference-Making. *Noûs*, 44(1), 80-111.

Kripke, S. (1972). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kronz, F., & Tiehen, J. (2002). Emergence and Quantum Mechanics, *Philosophy of Science* 69: 324–347.

Kusch, I., & Markus, M. (1996). Mollusc Shell Pigmentation: Cellular Automaton Simulations and Evidence for Undecidability. *Journal of Theoretical Biology*, 178(3), 333–340.

Ladyman, J., & Ross, D. (2010). *Every thing must go: Metaphysics naturalized*. Oxford: Oxford University Press.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago Press.

Landsberg, P. T. (1999). *Seeking ultimates: An intuitive guide to physics*. Boca Raton: CPC Press.

Laughlin, R. B., & Pines, D. (2000). The theory of everything. *Proceedings of the National Academy of Sciences*, 97(1), 28-31.

Laughlin, R. B. (2008). *A different universe: Reinventing physics from the bottom down*. New York: Basic books.

Lawhead, J. (2015). Self-organization, emergence, and constraint in complex natural systems. arXiv preprint, arXiv:1502.01476.

Lebowitz, J. L. (1999). Statistical mechanics: A selective review of two central issues. *Reviews of Modern Physics*, 71(2), S346.

Lewes, G. H. (1877). *Problems of life and mind*. Second series, or *The Physical Basis of Mind*. London: Trübner & Company.

Lewis, D. (1987). *Philosophical Papers*. Volume II, Oxford: Oxford University Press.

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76-80.

Liu, C. (1999). Explaining the emergence of cooperative phenomena. *Philosophy of Science*, 66, S92-S106.

Lizier, J. T. (2013). *The Local Information Dynamics of Distributed Computation in Complex Systems*. Berlin, Springer.

Lloyd, S. (2001). Measures of complexity: a nonexhaustive list. *IEEE Control Systems Magazine*, 21(4), 7-8.

Luisi, P. L. (2002). Emergence in Chemistry: Chemistry as the Embodiment of Emergence. *Foundations of Chemistry* 4: 183–200.

MacKay, R. S. (2008). Nonlinearity in complexity science. *Nonlinearity*, 21(12), 273-281.

Manafu, A. (2011). *Emergence and Reduction in Science. A Case Study.* Electronic Thesis and Dissertation Repository. 345.

Manafu, A. (2014). Concepts of Emergence in Chemistry. In J.P. Llored, (Ed.) *The philosophy of chemistry: practices, methodologies, and concepts*. Cambridge: Cambridge Scholars Publishing.

Manukyan, L., Montandon, S. A., Fofonjka, A., Smirnov, S., & Milinkovitch, M. C. (2017). A living mesoscopic cellular automaton made of skin scales. *Nature*, 544(7649), 173-179.

Marino, A. (2017). Riferimento singolare. *APhEx* 16, 1-18.

Marmodoro, A., & Mayr, E. (2019). *Metaphysics: An Introduction to Contemporary Debates and Their History*. Oxford: Oxford University Press.

Marsh, L., & Onof, C. (2008). Stigmergic epistemology, stigmergic cognition. *Cognitive Systems Research*, 9(1-2), 136–149.

Martin, C. B. (1994). Dispositions and conditionals. *The Philosophical Quarterly* (1950-), 44(174), 1-8.

Mattingly, J. (2013). Emergence of spacetime in stochastic gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3), 329–337.

McIntyre, L. (1998). Complexity: A philosopher's reflections. Complexity, 3, 26–32.

McLaughlin, B. P. (1992). The rise and fall of British Emergentism. In A. Beckerman, H. Flohr, & J. Kim (Eds.), *Emergence or reduction? Essays on the prospects of nonreductive physicalism* (pp. 49–93). Berlin: Walter de Gruyter.

McLaughlin, B. P. (1997). Emergence and supervenience. *Intellectica*, 25, 25–43.

Menzies, P. (2013). Mental Causation in the Physical World. In S.C. Gibb, E.J. Lowe, & R.D. Ingthorsson (Eds.). *Mental causation and ontology* (58-87). Oxford: Oxford University Press.

Merricks, T. (2003). *Objects and Persons*. Oxford: Clarendon Press.

McKaughan, D.J. (2013). Brute facts. In R. Fastiggi (Ed.), *New Catholic Encyclopedia* (p. 189). Boston: Cengage Learning.

Mill, J. S. (1844). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation*. London: John Parker, West Strand.

Mitchell, M. (2011). *Complexity: a guided tour*. New York: Oxford University Press.

Morgan, C. L. (1913). *Spencer's Philosophy of Science*. Oxford: Clarendon Press.

Morowitz, H. J. (2002). *The emergence of everything: How the world became complex*. Oford: Oxford University Press.

Muller, F.A. & S. Saunders. (2008). Discerning fermions. *The British Journal for the Philosophy of Science*, 59(3), 499–548

Nagel, E. (1949). The meaning of reduction in the natural sciences. In Stauffer, R. P. (Ed.) *Science and civilization* (pp. 97-135). Madison: Wisconsin UP.

Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. New York: Harcourt, Brace & World.

Nagel, E. (1970). Issues in the logic of reductive explanations. In M.A. Bedau, & P.E. Humphreys (2008). *Emergence: Contemporary readings in philosophy and science* (pp. 359-373). MIT press.

Newmann, D. V. (1996). Emergence and strange attractors. *Philosophy of Science*, 63, 245–261.

Nickles, T. (1973). Two concepts of intertheoretic reduction. *The Journal of Philosophy*, 70(7), 181-201.

Nicolis, S., & Deneubourg, J. (1999). Emerging Patterns and Food Recruitment in Ants: an Analytical Study. *Journal of Theoretical Biology*, 198(4), 575–592.

Norton, J. (2015). Weak discernibility and relations between quanta. *Philosophy of Science*, 82(5), 1188-1199.

O'Connor, T. (1994). Emergent properties. *American Philosophical Quarterly*, 31:91–104.

O'Connor, T. & Hong Yu Wong (2005). The Metaphysics of Emergence. *Nous*, 39:658–678.

Onnis, E. (2019) (Ed.). Parti, insiemi e sistemi. Il concetto di emergenza in filosofia. *Philosophy Kitchen*, 11.

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In Feigl, H., Scriven, M., & Maxwell, G. (Eds.), *Minnesota Studies in the Philosophy of Science*, Volume II (3–36). Minneapolis: University of Minnesota Press.

Papineau, D. (2001). The rise of physicalism. Reprinted In M.W.F Stone, & J. Wolff (Eds.). (2013). *Proper Ambition of Science* (pp. 182-216). Routledge.

Parham, P., & Janeway, C. (2009). *The Immune System*. London: Garland Science.

Paternoster, A. (2001). *Linguaggio e visione*. Pisa: ETS.

Pattee, H.H. (1972). Laws and Constraints, Symbols and Languages. In C.H.Waddington (ed.) *Towards a Theoretical Biology 4* (248-258), Edinburgh: Edinburgh University Press.

Pavarini, E., Koch, E., & Schollwöck, U. (Eds.). (2013). *Emergent Phenomena in Correlated Matter: Autumn School Organized by the Forschungszentrum Jülich and the German Research School for Simulation Sciences at Forschungszentrum Jülich 23-27 September*

2013; Lecture Notes of the Autumn School Correlated Electrons 2013 (Vol. 3). Forschungszentrum Jülich.

Pettit, P. (1993). A definition of physicalism. *Analysis*, 53(4), 213-223.

Pines, D. (2000). Quantum protectorates in the cuprate superconductors. *Physica C: Superconductivity* (341–348), 59–62.

Poe, E.A. (1903). *The Works of Edgar Allan Poe*, *the Raven Edition*, Vol. 5. New York: P. F. Collier and Son

Polani, D. (2013). Foundations and formalizations of self-organization. In Prokopenko, M. (Ed.). Advances in applied self-organizing systems (pp. 23-42). Springer, London.

Polanyi, M. (1968). Life's irreducible structure. *Science*, 160, 1308B1312.

Portugali, J. (2011). *Complexity, cognition and the city.* Springer Science & Business Media.

Post, H. (1963). Individuality and Physics, *The Listener*, 70: 534–537.

Poundstone, W. (1985). *The recursive universe: Cosmic complexity and the limits of scientific knowledge*. New York: Morrow.

Prokopenko, M., Boschetti, F., & Ryan, A. J. (2009). An information-theoretic primer on complexity, self-organization, and emergence. *Complexity*, 15(1), 11–28.

Quine, W. V. O. (1980). From a logical point of view: 9 logico-philosophical essays (Vol. 9). Harvard University Press.

Regenmortel, M. H. V. V. (2004). Reductionism and complexity in molecular biology. *EMBO Reports*, 5(11), 1016–1020.

Regenmortel M. H. V. V. (2004) Biological complexity emerges from the ashes of genetic reductionism. *Journal of Molecular Recognition,* 17: 145–148

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *Proceedings of the 14th annual conference on Computer graphics and interactive techniques* - SIGGRAPH 87.

Resnick, M. (1997). *Turtles, termites, and traffic jams: explorations in massively parallel microworlds*. Cambridge, Mass.: MIT Press.

Rosch, E. (1973). On the Internal structure of perceptual and semantic categories. In T. E. Moore (Ed.). *Cognitive development and the acquisition of language* (111–144). New York: Academic Press.

Rosen, G. (2010). Metaphysical Dependence: Grounding and Reduction. In R. Hale and A. Hoffman (eds.), *Modality: Metaphysics, Logic, and Epistemology*, Oxford: Oxford University Press, pp. 109–136.

Ross, D. (2000). Rainforest realism: A Dennettian theory of existence. In D. Ross, A. Brook., & D. Thompson (Eds.). (2000). *Dennett's philosophy: a comprehensive assessment* (147-168). MIT Press.

Sartenaer, O. (2018). Flat Emergence. *Pacific Philosophical Quarterly*, 99, 225-250.

Scerri, E. (2008). Reduction and Emergence in Chemistry - Two Recent Approches. *Collected Papers on Philosophy of Chemistry*, 71–88.

Schaffer, J. (2004). Two conceptions of sparse properties. *Pacific Philosophical Quarterly*, 85: 92–102.

Schaffer, J. (2004). Causes need not be physically connected to their effects: The case for negative causation. In C.R. Hitchcock (Ed.) (2004). *Contemporary debates in philosophy of science* (197-216). London: Blackwell.

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

Schaffer, J. (2008). Causation and laws of nature: In Sider, T., Hawthorne, J., & Zimmerman, D. W. (Eds.). *Contemporary debates in metaphysics*. John Wiley & Sons.

Schaffer, J. (2009). On What Grounds What. In Chalmers, D., Manley, D., & Wasserman, R. (Eds.). (2009). *Metametaphysics: new essays on the foundations of ontology*. Oxford University Press.

Schaffner, K. F. (1977). Reduction, reductionism, values, and progress in the biomedical sciences. *Logic, laws, and life*, 6, 143–171.

Schaffner, K. F. (1993). *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.

Schaffner, K. F. (2006). Reduction: the Cheshire cat problem and a return to roots. *Synthese*, 151(3), 377-402.

Schaffner, K. F. (2012). Ernest Nagel and reduction. *The Journal of Philosophy*, 109(8/9), 534–565.

Seager, W. (2014). Why Physicalism?. *Mind and Matter*, 12(2), 143-195.

Searle, J. R. (1958). Proper names. *Mind*, 67(266), 166-173.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Shoemaker, S. (2007). *Physical realization*. Oxford: Oxford University Press.

Sklar, L. (1967). Types of inter-theoretic reduction. *The British Journal for the Philosophy of Science*, 18, 109–124.

Sklar, L. (1993). *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

Silberstein, M. (1999). The search for ontological emergence. *Philosophical Quarterly* 49: 182–200.

Silberstein, M. (2002). Reduction, emergence and explanation. In P. Machamer, & M. Silberstein (Eds.). (2008). *The Blackwell guide to the philosophy of science* (80-107). Hoboken, New Jersey: John Wiley & Sons.

Silberstein, M. (2011). Metaphysics or science: The battle for the soul of philosophy of mind. *Philosophical Psychology* 24: 561–57

Silberstein, M. (2012). Emergence and reduction in context: Philosophy of science and/or analytic metaphysics. *Metascience* 21(3): 627–641.

Silberstein, M., & McGeever, J. (1999). The search for ontological emergence. *The Philosophical Quarterly*, 49(195), 201-214.

Siganos, G., Tauro, S. L., & Faloutsos, M. (2006). Jellyfish: A conceptual model for the as Internet topology. *Journal of Communications and Networks*, 8(3), 339–350.

Sonntag, R. E., Borgnakke, C., Van Wylen, G. J., & Van Wyk, S. (1998). *Fundamentals of thermodynamics.* New York: Wiley.

Symons, J. (2018). Brute facts about emergence. In E. Vintiadis, & C. Mekios (Eds.). (2018). *Brute Facts* (pp.). Oxford University Press.

Stöckler, M. (1991). A Short History of Emergence and Reductionism. In Agazzi, E. (Ed.). (2012). *The Problem of Reductionism in Science*. Berlin: Springer.

Stoljar, D. (2010). *Physicalism*, Abingdon: Routledge.

Strawson, P. F. (2002). *Individuals*. Abingdon: Routledge.

Subramanian, N., Torabi-Parizi, P., Gottschalk, R. A., Germain, R. N., & Dutta, B. (2015). Network representations of immune system complexity. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 7(1), 13-38.

Sumpter, D. J., Mann, R. P., & Perna, A. (2012). The modelling cycle for collective animal behaviour. *Interface Focus*, 2(6), 764-773.

Theraulaz, G., & Bonabeau, E. (1999). A Brief History of Stigmergy. *Artificial life*. 5, 97-116.

Trible, W. & Kronauer, D. J. C. (2017) Caste development and evolution in ants: it's all about size. *Journal of Experimental Biology* 220, 53-62.

Tripet, F., & Nonacs, P. (2004). Foraging for Work and Age-Based Polyethism: The Roles of Age and Previous Experience on Task Choice in Ants. *Ethology*, 110(11), 863-877.

Wallace, A.R. (1870). *Contributions to the Theory of Natural Selection: A Series of Essays*. London: Macmillan and Co.

Van Brakel, J. (2000). *The Philosophy of Chemistry*. Leuven University Press.

VandeWall, H. (2007). Why water is not H2O, and other critiques of essentialist ontology from the philosophy of chemistry. *Philosophy of Science*, 74(5), 906-919.

Van Gulick, R. (2001). Reduction, emergence and other recent options on the mind/body problem. A philosophic overview. *Journal of Consciousness Studies*, 8(9-10), 1-34.

van Riel, Raphael and Van Gulick, Robert, "Scientific Reduction", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), forthcoming URL = https://plato.stanford.edu/archives/sum2018/entries/scientific-reduction/

Vintiadis, E., & Mekios, C. (Eds.). (2018). *Brute facts*. Oxford University Press.

Violi, P. (1997). *Significato ed esperienza*. Milano: Bompiani.

Weaver, W. (1948). "Science and Complexity". *American Scientist*, 36: 536-547.

Weisberg M. (2006) Water is Not H2O. In D. Baird, E. Scerri, L. McIntyre (Eds.). *Philosophy Of Chemistry. Boston Studies in the Philosophy of Science* (337-345), vol 242. Dordrecht: Springer.

Weisstein, E. W. "Rule 30". From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/Rule30.html

West, G. B. (2017). *Scale: the universal laws of growth, innovation, sustainability, and the pace of life in organisms, cities, economies, and companies*. London: Penguin.

Wilson, J. (1999). How superduper does a physicalist supervenience need to be?. *The Philosophical Quarterly*, 49(194), 33-52.

Wilson, J. (2005). Supervenience-based formulations of physicalism. *Nous*, 39(3), 426-459.

Wilson, J. (2013). Nonlinearity and Metaphysical emergence. In S. Mumford & M. Tugby (Eds.), *Metaphysics and science*. Oxford: Oxford University Press.

Wilson, J. (2014). No work for a theory of grounding. *Inquiry*, 57(5-6), 535-579.

Wilson, J. (2016). Metaphysical emergence: Weak and Strong. In T. Bigaj, & C. Wuthrich, (Eds.), *Metaphysics in Contemporary Physics* (pp. 251–306). Leiden, Brill.

Wilson, J. (2019). *Metaphysical Emergence*. Oxford: Oxford University Press.

Wilson, R. A. (1999). *Species: New interdisciplinary essays*. Cambridge: the MIT Press.

Wimsatt, W. (1976). Reductive explanation: A functional account. In *Proceedings of the 1974 Biennial Meeting Philosophy of Science Association* 1974 (pp. 671–710). Berlin: Springer.

Wimsatt, W. (1986). Developmental Constraints, Generative Entrenchment, and the Innate-Acquired Distinction. In W. Bechtel (ed.), *Integrating Scientific Disciplines*. Dordrecht: Martinus Nijhoff.

Wimsatt, W. C. (1994). The ontology of complex systems: levels of organization, perspectives, and causal thickets. *Canadian Journal of Philosophy*, 24, 207-274.

Wimsatt, W. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science* 64 (4): 372–84.

Wimsatt, W. (2000). Emergence as Non-Aggregativity and the Biases of Reductionisms. *Foundations of Science* 5: 269.

Wimsatt, W., and Sarkar, S. (2006). "Reductionism". In Pfeifer, J., and Sarkar, S. (Eds.), *Philosophy of Science: An Encyclopedia*. New York: Routledge. pp. 696–703.

Wittgenstein, L. (2009). *Philosophical investigations*. Hoboken: John Wiley & Sons.

Wolfram, S. (1984a). University and complexity in cellular automata. *Physica*, D 10, 1–35

Wolfram, S. (1984b). Cellular automata as models of complexity. *Nature*, 311(5985), 419–424.

Wolfram, S. (1985). Undecidability and intractability in theoretical physics. Physical Review Letters, 54, 735–738.

Wolfram, S. (1986). "Tables of Cellular Automaton Properties", in *Theory and Applications of Cellular Automata*. World Scientific Publishing Co. Ltd. 485–557.

Wolfram, S. (2002). A new kind of science. Champaign, IL: Wolfram Media.

Wordsworth, W. (1802/1992). My heart leaps up when I behold. In N. Roe (Ed.), *William Wordsworth selected poetry* (p. 172). New York: Penguin Books.

Yates, David (2016). Demystifying Emergence. *Ergo, an Open Access Journal of Philosophy* 3 (31), 809–844

Zhang, S. (2004). To see a world in a grain of sand. In J. D. Barrow, P. C. W. Davies, & C. L. Harper (Eds.) *Science and ultimate reality: quantum theory, cosmology, and complexity*. Cambridge: Cambridge University Press