AN EMERGENT COSMOS: AN EXPLORATION AND DEFENSE OF THE CONCEPT OF EMERGENCE

MARTEN KAAS, M.Sc., B.Sc. and B.A.

An Emergent Cosmos: An Exploration and Defense of the Concept of Emergence

by Marten Kaas, M.Sc., B.Sc. and B.A.

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree

Master of Arts

MASTER OF ARTS (2018) McMaster University
(Philosophy) Hamilton, Ontario

TITLE: An emergent cosmos: an exploration and defense of the concept of emergence

AUTHOR: Marten Kaas, M.Sc., B.Sc. and B.A.

SUPERVISOR: Dr. Sandra Lapointe

NUMBER OF PAGES: vii, 114

Abstract

The concept of emergence stands in need of an update, and I propose that ontologically emergent phenomena are characterized by four necessary features: relationality, novelty, irreducibility and broken symmetry. ‘Emergence’ is a useful term to denote the varied qualitative changes that spontaneously arise as the scale and complexity of related phenomena increases. Moreover, emergent phenomena share a unique relationship with the phenomena from which they emerge, namely the emergent relation. This relation is distinct from other types of relations (i.e., identity, composition, supervenience, etc.) and moreover is not beset by the problems of causal exclusion or downward causation. Lastly, I advance this account of emergence partly as an empirical hypothesis. The epistemic resources in dynamical systems theory are uniquely suited to describe the evolution of systems that manifest emergent phenomena. This is primarily because features like novelty and broken symmetry can be given mathematically precise descriptions in dynamical systems terms. The advantage of this updated concept of emergence is its compatibility with ideas of explanation, prediction and reduction.

Acknowledgements

I would like to offer my deepest thanks to Dr. Sandra Lapointe for such an amazing opportunity and her faith in my abilities as a scholar. Your guidance and support have helped me become a better student and philosopher.

I would also like to thank Dr. Diane Enns for her mentorship during my undergraduate studies and continued support throughout graduate school as well as Dr. Richard Arthur for his insightful comments and support.

A big thank you to my peers in the department and those students in my cohort for their continued friendship and guidance. Special shout out to the Social Committee.

I would like to thank Nashwa, my best friend, for supporting me through this journey and for helping me find my way when I am lost.

Lastly I would like to thank my family from whom I have emerged.

Table of Contents

Introduction

Our universe is simply fascinating. The impetus for this thesis project stems from a profound wonder I have of the natural world and the phenomena within it. Living organisms push back against the corrosive and erosive forces of a universe tending towards thermodynamic equilibrium. The Earth itself heaves and groans as gravity draws matter towards its center and quantum mechanical forces push back. How are we to understand the relatedness and connectedness of these phenomena? The concept of “emergence” was introduced by philosophers and scientists to describe related phenomena that, intuitively, appear to be radically different. The relation between mind and the body for example has been of interest for millennia. What is the relationship between mental phenomena, like beliefs, intentions, creativity, intelligence and desires, and biological phenomena, like neurons, brains and nervous systems?

Emergence is a concept closely linked to concepts like explanation, prediction, reduction and construction, all of which have important roles in the history of philosophy of mind as well as philosophy of science in general. One general aim of this thesis is to untangle the relationship between these concepts in the context of an investigation of emergence and emergent phenomena. The second general aim is to put forward and defend a coherent account of ontological emergence. The third general aim of this thesis is to motivate the use of the term ‘emergence’ in the context of scientific as well philosophical investigation. That is, the account of emergence advanced in this thesis is intended to be both philosophical as well as empirical. Ultimately, I see myself as advancing a concept that has the potential to parsimoniously unite disparate phenomena. Indeed I believe such a concept is necessary given the rapid and surprising advances in the fields of artificial intelligence and machine learning.

Chapter 1

An Old Rivalry: Reduction and Emergence

The concept of emergence and emergent phenomena has a rich history that stretches over one hundred years. Introduced in the late 19th century and early 20th century, the concept of emergence enjoyed a rise in popularity among philosophers and scientists. Emergence owes its prominence to the scholars who belong to a group commonly referred to as the early British Emergentists, and include figures such as John Stuart Mill, C. Lloyd Morgan, Charlie Broad and Samuel Alexander.[1] Reacting against the mechanistic reductionism that had come to dominate the sciences even before the discovery of Leibniz's/Newton's calculus, the British Emergentists insisted that not all phenomena are understandable in such reductive terms. They sought to understand what they thought were genuinely novel phenomena that appeared to resist a relatively straightforward reduction to mechanistic terms. Leibniz's famous mill argument and the comments he makes in Section 17 of the *Monadology* highlight the way in which mechanistic reduction was commonly understood before the concept of emergence was concisely articulated. Leibniz proposes the following thought experiment. Imagine that we constructed a machine that could think, feel and perceive in the way that human beings, for example, think, feel and perceive, and imagine further that such a machine was enlarged to the dimensions of a

[1] McLaughlin 2008, 19.

mill so that one could walk around inside of it.[2] Where would we locate "perception" or "thinking?" According to Leibniz, these are not to be found in the "shapes and motions" of the parts of our machine.[3]

Despite the paradigm shifting success of Newtonian mechanics, it was becoming apparent to some philosophers that there existed phenomena that resisted description simply in terms of sums of vectors of motion of their constituents, i.e., they resisted description in wholly mechanistic term. The seeming irreducibility of a variety of complex phenomena which included chemistry, non-linear systems and life itself, coupled with the introduction of the idea of Darwinian evolution, prompted some scholars to search for a framework under which such phenomena could be understood. Moreover, given the broadly physicalist/materialist[4] commitments held by philosophers and scientists at the turn of the 20th century, proposed conceptual frameworks enlisted to explain these complex phenomena had to shed the aura of unintelligibility that pervaded explanations of biological and mental phenomena at the time. It was frequent in the early 20th century to invoke Cartesian souls or the presence of entelechies to understand a suite of biological and mental phenomena including reproduction and consciousness. Enter the early British Emergentists.

> We live in a world in which there seems to be an orderly sequence of events. It is the business of science, and of a philosophy which keeps in touch with science, to describe

---

[2] Arthur 2014, 70-71.
[3] Ibid., 70.
[4] For our purposes I will treat physicalism and materialism as synonyms referring to a metaphysical position that there is nothing more than, or nothing but physical objects, events and their properties. In its strongest form, this position entails a commitment to eliminative physicalism which would deny the existence of nonphysical objects, events or properties such as mental objects, events or properties. However the physicalist need not take such a strong position, since it is possible to hold that although there exist nonphysical objects, events or properties, these are nevertheless reducible to fundamentally physical objects, events or properties.

> the course of events in this or that instance of their occurrence, and to discover the plan on which they proceed...But the orderly sequence, historically viewed, appears to present, from time to time, something genuinely new. Under what I here call emergent evolution stress is laid on this incoming of the new. Salient examples are afforded to the advent of life, in the advent of mind, and in the advent of reflective thought.[5]

The radical novelty of certain phenomena prompted scholars to protest against a "mechanistic dogma" that permeated philosophy and the sciences.[6] According to Morgan, the essential feature of this mechanistic interpretation of phenomena in the world "is that it is in terms of resultant effects only, calculable by algebraic summation."[7] Such a worldview neglects to account for the "something more" necessary for the appearance of emergent phenomena.[8] This idea of irreducibility, that emergent phenomena are something "more than" the sum of their constituents is characteristic of most, if not all, early British Emergentist views. For example, writing on the mind and mental phenomena, Alexander notes that "while mental process is *also* neural, it is not *merely* neural" and is therefore an emergent phenomena.[9] Mind requires something more than a neural assemblage, something more than certain vital processes. Of course Alexander was only able to speculate on what that "something more" might be, writing that "mind requires...a collocation of conditions," perhaps located in the nervous system, that together, somehow, give rise to the newness of mind.[10]

The idea of irreducibility, that some phenomena could not be understood simply on the basis of an examination of their constituent components, is a characteristic feature of emergence as it was articulated by the early British Emergentists. Although they also held a

---

[5] Morgan 1923, 1.
[6] Ibid., 8.
[7] Ibid.
[8] Ibid.
[9] Alexander 1920, 6.
[10] Ibid.

broadly physicalist worldview, i.e., that everything is made of matter and all motion is determined by Newtonian mechanics, the British Emergentists resisted the tendency to reductively explain and eliminate certain phenomena via their decomposition into constituent parts. The concept of emergence was advanced as one of the first of many concepts under the doctrine of nonreductive physicalism in an attempt to reconcile strict physicalism with outright dualisms. Unsurprisingly then, British Emergentists endorsed a hierarchical conception of phenomena determined by the organizational complexity of matter that begins, at bottom, with the physical and progresses through the chemical, biological and finally psychological.[11] Alexander for example asserts that the obviously distinguishable levels of existence include "motions, matter as physical (or mechanical), matter with secondary qualities, life, [and] mind."[12] He elaborates on the idea of the emergence of a new quality using the paradigmatic mind-body relation.

> Physical and chemical processes of a certain complexity have the quality of life. The new quality life emerges with this constellation of such processes, and therefore life is at once a physico-chemical complex and is not merely physical and chemical, for these terms do not sufficiently characterise the new complex which in the course and order of time has been generated out of them...The higher quality emerges from the lower level of existence and has its roots therein, but it emerges therefrom, and it does not belong to that lower level, but constitutes its possessor a new order of existent with its special laws of behaviour. 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. It admits no explanation.[13]

[11] McLaughlin 2008, 20.
[12] Alexander 1920, 52.
[13] Ibid., 46-47.

Morgan similarly argues that events at the level of life are above events at the level of matter, with events at the level of mind being higher still. Furthermore, Morgan explains the sense in which some phenomena are "higher" than others.

> When two or more kinds of events, such as I spoke of before [i.e., mental, biological and physical events] as A, B and C, co-exist on one complex system in such wise that the C kind involves the co-existence of B, and B in like manner involves A, whereas the A-kind does not involve the co-existence of B, nor B that of C, we may speak of C, as, in this sense, higher than B, and B than A. Thus, for emergent evolution [i.e., the existence of emergent phenomena], conscious events at level C (mind) involve specific physiological events at level B (life), and these involve specific physico-chemical events at level A (matter). No C without B, and no B without A. No mind without life; and no life without "a physical basis."[14]

Both Alexander and Morgan illustrate how proponents of emergence in the early 20th century believed that ontologically novel phenomena, i.e., phenomena that constitute a new order of existence, arise naturally as a result of the temporal evolution of the universe. Emergent phenomena are irreducible to the phenomena from which they emerge, but are nevertheless inextricably related to the phenomena from which they emerge.

Early British Emergentists, in addition to their claims of ontological irreducibility and novelty of emergent phenomena, also maintained the epistemic irreducibility and unexplainability/unpredictability of emergent phenomena. Rejecting the Laplacean coupling of causal determinism to explanation and prediction,[15] British Emergentists insisted that emergent phenomena are fundamentally unpredictable. Morgan proposes a thought experiment in which

---

[14] Morgan 1923, 15.

[15] Given that Newton's laws of motion are time symmetric, Pierre Simon Laplace articulated a thought experiment to demonstrate the causal determinism implied by Newtonian mechanics. In his thought experiment, a superintelligent entity capable of knowing the location and momentum of all particles in the universe would be able "see" with certainty the past and future of the universe.

we are to imagine ourselves as sentient beings living in the "fire-mist" of the extremely young universe.[16] Supposing that this is a time prior to the evolution of crystalline solids (perhaps only gaseous hydrogen and minimal amounts of gaseous helium exists), Morgan asks: "Could we then, on the basis of the fullest possible experience of our fire-mist world, foretell the forms that crystalline synthesis would assume in the not-yet of the future?"[17] He answers in the negative, it is not possible for us as sentient beings, even using the scientific method, to predict facts the like of which have not yet "swum into the ken of experience."[18] Morgan explicitly rejects mechanism and the causal determinism Newtonian mechanics implies.

> I [Morgan] hold that all scientific explanation is after the event, and that all scientific prediction is of like events under like conditions. But surely, it may be urged, an adequate knowledge of the constitution of nature would enable us to predict any event no matter how novel or how far removed from us in future time. In a sense this is true enough—but only in the sense that the supposed adequate knowledge embraces the constitution of nature *when it is finished*—if it ever gets finished for human understanding to grasp. In the case I have supposed, the order of nature as an evolutionary product was still in the making and had not reached the critical moment of crystallization.[19]

Morgan is adamant that no amount of knowledge concerning the present or past state of the universe is sufficient to predict what sort of emergent phenomena may arise as a result of the temporal evolution of our universe. It is not entirely clear however whether Morgan believes that emergent phenomena are unexplainable. His concession that we may have adequate knowledge of a phenomenon only after that phenomenon has arisen (e.g., we have adequate knowledge of crystalline solids because they have "evolved" already and we are able to study

[16] Morgan 1912, 148.
[17] Ibid.
[18] Ibid., 149.
[19] Ibid.

them) seems to indicate that despite their unpredictability, emergent phenomena are nevertheless understandable. In short, we cannot predict that which is simply outside the realm of experience but we can come to understand, and perhaps explain after the fact, that which is within the realm of experience. Instead of unpredictability, Alexander stresses the unexplainability of emergent phenomena. It is just “a matter of observed empirical fact” that out of the temporal evolution of the universe and the complexity of the motion of the particles within the universe that new qualities, emergent qualities, appear.[20] The emergence of mental phenomena from biological phenomena for example, is something that ought to just be accepted as a “brute empirical fact” that “admits of no explanation.”[21]

Reduction, Explanation and Prediction

Unfortunately for the British Emergentists, interest in the concept of emergence and emergent phenomena sharply dropped following the quantum mechanical and genetic revolutions of the mid-20th century. Advanced, in part, as an empirical hypothesis, it seemed that the evidence supported a primarily reductionist view of phenomena in the universe. That is, in contrast to some of the core doctrines of emergence (e.g. the irreducibility, unexplainability and unpredictability of certain phenomena), it appeared that chemical and biological phenomena (among many other phenomena) were reducible, explainable and predictable: the reductionist worldview, that properties of systems as a whole could be understood on the basis of their constituents, appeared to be empirically vindicated. In a radical departure from narrow

[20] Alexander 1920, 45.
[21] Ibid., 46-47.

Newtonian mechanics, quantum mechanics broadened our ideas of "mechanism" and as a result our ideas of reductive explanation.

> The members of the British Emergentist tradition were perfectly correct in claiming that the product of two chemical reactants is in no sense the sum of what would have been the effect of each reactant had it acted alone. Chemical processes indeed produce heteropathic or emergent effects; and chemical laws are indeed heteropathic or emergent...But that chemistry is emergent in the sense in question poses no problem for reductive materialism...Quantum mechanics reductively explains chemistry, but without appeal to additive or even linear compositional principles, and without the postulation of new irreducible higher-level forces. Moreover, quantum mechanics has led to the development of molecular biology, and the successes of this discipline has virtually eradicated any sort of vitalism from biology.[22]

Emergence was no longer considered a serious philosophical or scientific concept since it appeared that phenomena in the world simply were just the resultants, albeit a complex kind of resultant, of the fundamental constituents of the phenomena under scrutiny. Tied as it was to advances and discoveries in the sciences, the concept of emergence simply could not withstand mounting empirical evidence that supported a reductionist worldview.

Throughout the mid-20th century interest in emergence waned as interest in reduction surged. It was during this time that Carl Hempel and Paul Oppenheim (hereafter H&O) advanced their canonical account of scientific explanation connecting logical deduction, i.e., reduction, to explanation and prediction. In their paper *Studies in the Logic of Explanation*, H&O note that their formal analysis of scientific explanation applies just as well to scientific prediction. On their view, empirical science is concerned not only with the question of "what?" but also the question of "why?" in regards to phenomena within the world.[23] To give a scientific explanation of a

[22] McLaughlin 2008, 88.
[23] Hempel and Oppenheim 1948, 9.

particular phenomenon is therefore to satisfy the “what?” and “why?” questions. When a mercury thermometer is placed in hot water there is first a temporary drop in the column of mercury followed by a swift rise: why did this particular phenomenon occur?[24] The answer is that the increase of temperature first affects the glass of the thermometer, expanding it, resulting in a drop in the level of mercury. Only then does the temperature change reach the mercury itself, resulting in a much greater expansion that causes the observed rise in the column of mercury. H&O note that in answering the “why?” question in regards to the particular phenomenon of the fall and rise of the mercury, their account consists of statements of two kinds. Specifically, their account consists of statements about antecedent or initial conditions and also statements about certain general laws.[25] The result of this examination of a relatively simple phenomenon, specifically keeping in mind the questions of “why?,” provides a model of what it is to be able to scientifically explain a particular phenomenon.

Keeping in mind the sketch of scientific explanation above, H&O proceed to draw out certain general characteristics of scientific explanation. First, explanation can be divided into two major constituents: the thing to be explained and the thing by which we do the explaining. The thing to be explained is the “explanandum” by which H&O “understand the sentence describing the phenomenon to be explained (not that phenomenon).”[26] The thing by which we do the explaining is the “explanans,” understood as “the class of those sentence which are adduced to account for the phenomenon.”[27] The explanans is further divided into two subclasses previously identified, namely initial conditions and general laws. Given their focus on scientific or, what

[24] Ibid.
[25] Ibid., 9-10.
[26] Ibid., 10.
[27] Ibid.

amounts to the same as H&O see it, "causal explanation" in particular, this conception of explanation requires that certain logical and empirical conditions of accuracy be met. These conditions are as follows.

> (R1) The explanandum must be a logical consequence of the explanans; in other words, the explanandum must be logically deducible from the information contained in the explanans; for otherwise, the explanans would not constitute adequate grounds for the explanandum.
> (R2) The explanans must contain general laws, and these must actually be required for the derivation of the explanandum...
> (R3) The explanans must have empirical content; that is, it must be capable, at least in principle, of test by experiment or observation. This condition is implicit in (R1); for since the explanandum is assumed to describe some empirical phenomenon, it follows from (R1) that the explanans entails at least on consequence of empirical character, and this fact confers upon it testability and empirical content...
> (R4) The sentences constituting the explanans must be true...[28]

Thus we have arrived, at least according to H&O, at a preliminary account of the essential characteristics of scientific explanation. This is the earliest formulation of what will come to be known as the deductive-nomological (DN) model of explanation. Deductive because H&O insist that one must be able to logically deduce the explanandum given the explanans, and nomological because their model of explanation requires the inclusion of general law statements connecting the phenomena under study.

The crucial point here is that H&O maintain that this account of scientific explanation applies equally to scientific prediction. The only difference between an explanation and a prediction is a pragmatic one, to wit the temporal relation between the observation of a phenomenon and its proposed explanans. H&O argue that given the occurrence of a phenomenon described by the explanandum (E), and a suitable set of statements corresponding

[28] Ibid., 11.

to the relevant initial conditions (C1, C2,...Cy) and general laws (L1, L2,...Lz) is provided afterwards, we speak of an explanation, i.e., the phenomenon under scrutiny has been explained ex post facto.[29] If the initial conditions and general laws are given and E is derived prior to the observation of an occurrence of a phenomenon, we speak of prediction.[30] The phenomenon under scrutiny is said to have been predicted if, a priori, its occurrence is logically deduced from the explanans. Indeed it is H&O's first stipulated necessary condition of scientific explanation/prediction (R1) requiring that the explanandum be a logical consequence of the explanans that forces H&O to declare that a particular explanation is not fully adequate unless "its explanans, if taken account of in time, could have served as a basis for predicting the event in question."[31] The problem with this account of scientific explanation however, is that it is not at all clear that explanation and prediction actually enjoy such a logically symmetrical relationship. The relationships between the concepts of explanation, prediction, reduction and emergence will be a major focus of this thesis given their interconnectedness.

H&O's DN model of explanation had two major consequences: it relegated the concept of emergence to epistemology, and it coupled explanation and prediction together. Already a shunned concept, H&O stripped emergent phenomena of any ontological status they may have retained and placed the concept of emergence firmly in the realm of epistemology. Emergent phenomena, as phenomena that admit of no explanation and thus must be accepted as brute empirical facts, were simply outside the scope of H&O's DN model of explanation.

[29] Ibid., 12.
[30] Ibid.
[31] Ibid.

> Failure to realize that the question of predictability of a phenomenon cannot be significantly raised unless the theories available for the prediction have been specified has encouraged the misconception that certain phenomena have a mysterious quality of absolute unexplainability, and that their emergent status has to be accepted with "natural piety,"...emergence of a characteristic is not an ontological trait inherent in some phenomenon; rather it is indicative of the scope of our knowledge at a given time; thus it has no absolute, but a relative character; and what is emergent with respect to the theories available today may lose its emergent status tomorrow.[32]

Emergent phenomena, for H&O, are only labeled as such in lieu of a blatant admission of our ignorance of the initial conditions and general laws, those statements that figure in the explanans, required to deduce, and hence explain/predict, a certain phenomenon. H&O write with the obvious benefit of hindsight, but this does not detract from the fact that this is a fairly accurate description of the process by which early British Emergentists labelled certain phenomena emergent and others merely resultant. Reproduction and the passing of heritable traits from one generation to the next in biology, in fact life itself, were offered as examples of emergent phenomena without any awareness or understanding of molecular biology. These brute facts about the emergent phenomenon of life were no longer just that, brute facts, following the genetic revolution. Reproduction and heritable traits were explainable given the general laws of molecular biology.

Beyond stripping the ontological import of emergent phenomena, H&O's DN model of explanation logically coupled explanation and prediction. Moreover, this coupling was under a reductionist framework resulting in an inextricable link between the concepts of reduction and explanation/prediction. Explanations that are structured according to the DN model are said to reductively explain a phenomenon because the explanans entails the explanandum. The "general and unexceptional connections between specified characteristics of events" included in

[32] Ibid., 21.

the explanans (in the general law statements) determines, via logical deduction, the explanandum.[33] The explanandum can therefore be considered “reduced to” the explanans; the explanandum is nothing “more than” the explanans. Importantly, this is a primarily epistemological consequence of the DN model of explanation. That is, the reduction of the explanandum is an epistemological reduction and not necessarily an ontological one, hence the hybrid term ‘reductive explanation.’ By introducing a relationship of dependency between the explanans and explanandum Ernest Nagel argues that H&O’s DN model of explanation should be interpreted as implying something like “reduction:”

> ...certain relations of dependence between one set of distinctive traits of a given subject matter are allegedly explained by, and in some sense “reduced” to, assumptions concerning more inclusive relations of dependence between traits or processes not distinctive of (or unique to) that subject matter.[34]

Commonly cited examples of this kind of reduction include the explanation of the kinetic theory of heat in terms of Newtonian mechanics or chemical laws/interactions in terms of quantum mechanics. In general, modern science has strong reductive tendencies. A pervasive view of reality which we find in the work of philosophers and scientists of the mid-20th century (and which persists to this day) is one in which the “world is nothing but spatiotemporal arrangements of fundamental physical objects and properties.”[35] Everything else, all properties, objects and states of affairs, are merely rearrangements of the fundamental constituents of the universe. This is also a primarily epistemological consequence of the DN model of reductive explanation. This much is clear from the writing of H&O and other prominent proponents of the

[33] Ibid., 13.
[34] Nagel 2008, 360.
[35] Humphreys 2016, 1.

DN model of explanation who are explicit that by the terms ‘explanans’ and ‘explanandum’ they understand those sentences which are adduced to account for the phenomenon under study and the sentence describing the phenomenon, not the phenomenon itself, respectively.[36] Nagel identifies that this consideration, i.e., the distinction between the sentences describing a phenomenon and the phenomenon itself, applies equally to the term ‘reduction.’

> For strictly speaking, it is not phenomena which are deduced from other phenomena, but rather *statements* about phenomena from other statements. This is obvious if we remind ourselves that a given phenomenon can be subsumed under a variety of distinct descriptions, and that phenomena make no assertions or claims...Whatever else may be said about reductions in science, it is safe to say that they are commonly taken to be explanations, and I [Nagel] will so regard them. In consequence, I will assume that, like scientific explanations in general [especially those that conform to the DN model], every reduction can be construed as a series of statements, one of which is the conclusion (or the statement which is being reduced), while the others are the premises or reducing statements.[37]

Yet despite affirming that the terms ‘reduction,’ ‘explanation’ and ‘prediction’ are significant in the realm of epistemology, it is unclear whether the terms speak to any underlying ontology. H&O deny that the concept of emergence has any ontological force, but it is unclear whether the concepts of explanation/prediction carry an ontological component. One might argue that the most general or fundamental laws that can figure in an explanation/prediction, such as the laws that might appear in a Theory of Everything, would have ontological significance, but this is debatable. Consider that even if we consider H&O to be ontological reductivists, there is the further problem of identifying which version of reduction, e.g., eliminative or noneliminative reduction, they support. Each version of reduction has different ramifications on the ontological

[36] Hempel and Oppenheim 1948, 10.
[37] Nagel 2008, 360-361.

status of certain phenomena in the world ranging from their outright elimination from our ontology to their status as causally inert epiphenomena.

The Evidence Strikes Back: Emergence Episode II

The reductionist project, just like emergentist one, was put forward in part as an empirical thesis. And just as emergence was shunned in light of scientific progress in the early 20th century so too was reduction challenged and heavily scrutinized at the end of the 20th century. The tight logical coupling of explanation and prediction introduced by H&O in their DN model of explanation for example was heavily and immediately criticized. Heather Douglas notes that the relation via logical deduction between explanation and prediction came to be known as the "Symmetry Thesis" and drew considerable criticism throughout the late 1950's and 1960's.[38] Many philosophers argued that there are significant differences between explanation and prediction that precluded their sharing a symmetrical relationship. Michael Scriven for example claims that "explanation requires something 'more than' prediction" considering that predictions could be offered without the least bit of knowledge concerning why, the quintessentially H&O question of explanation, a particular phenomenon has occurred.[39] Consider that although a novice chess player might be able to make a prediction about the next move a grandmaster chess player might make, their prediction likely has no bearing on an explanation of why the grandmaster chose that particular move. Alternatively in Scriven's terms, to offer an explanation is to suggest that one possesses an understanding of a particular

[38] Douglas 2009, 448.
[39] Scriven 1962, 177.

phenomenon that is not required when simply forecasting a phenomenon.[40] A novice chess player, indeed even a person who has never seen a game of chess before, could predict, or attempt to predict, the next move in a game of chess simply by pointing to a square on the board and specifying what piece will occupy that square. This kind of scenario seems to suggest that there is a sense in which explanation and prediction are separate and distinct from one another.

Now a proponent of H&O's DN model of explanation would likely argue that the prediction described above, of where a particular chess piece will be during the next turn, is really no prediction at all on their account. If anything, it would be a prediction made out of ignorance of the explanans. A novice chess player may be ignorant of the initial conditions of the chess board and the general laws governing the behaviour of pieces on the board. Supposing that a person was supplied with the explanans however, the question remains: are explanation and prediction logically symmetrical? Phrased differently we might ask: are phenomena describable by the DN model of explanation causally determined by the explanans? I maintain that the answers to these questions are "no" and "not all phenomena." For one, consider that although scientists and philosophers alike tend to ignore its presence, time and the temporal evolution of our universe precludes the possibility of there ever being precisely identical initial conditions of a system under investigation. Yet this conclusion seems at odds with our everyday experience. Surely we do possess some amount of predictive capabilities given the success of human technology. Whole fields of inquiry, such as astrophysics and pharmacology, rely heavily on prediction. Solar and lunar eclipses can be predicted down to the minute and candidate drug compounds are assessed initially on their predicted bioactivity out of a pool of billions of

---

[40] Ibid.

chemical compounds. I am however merely echoing a point made by the early British Emergentists and subsequently ignored, namely that, contra H&O, time and temporal evolution are not simply pragmatic matters. Similarly, although the British Emergentists accepted that certain phenomena were reducible and certain other phenomena emergent, they maintained a strict distinction between scientific explanation and prediction. Explanations are offered ex post facto and scientific predictions are offered only for like events under like conditions, a weaker but ultimately more defensible position in comparison to the logically symmetrical relationship advocated by H&O.

These same early emergentist sentiments were recycled by philosophers such as Norwood Hanson who argues that H&O describe an ideal situation and relationship between explanation and prediction. Hanson maintains that "the history of science presents very few examples of disciplines wherein this optimal state of affairs [the logical symmetry of explanation and prediction] has actually been achieved," a fact that ought to raise questions about how H&O couple explanation and prediction together.[41] More importantly, despite the success of quantum mechanics in reductively explaining phenomena that were previously considered to be emergent, the premises of quantum theory itself appear to preclude even the theoretical possibility of reductively predicting, a la H&O's DN model, a particular phenomenon.

> Here [in quantum mechanics] the facts seem to run wholly against the schemata suggested by Hempel. It is perfectly true that, *given* any single quantum phenomenon P, P can be completely *explained* ex post facto; one can *understand* fully just what kind of event occurred, in terms of the well-established laws of the composite quantum theory of Jordan, Dirac, Heisenberg, and the later developments of Dyson, Schwinger, and Tomonaga...But it is, of course, the most fundamental feature of these laws that the *prediction* of such a phenomenon P is, as a matter of theoretical principle, quite

[41] Hanson 1959, 350.

> impossible...Put more directly, the only theory which explains (in any sense of "explain") quantum phenomena has built into it as a *notational rule of the system* the impossibility of predicting such phenomena.[42]

So quantum theory itself casts doubt on the intelligibility of the logical symmetry between explanation and prediction. It appears we have a reason to reject H&O's claim that an explanation is not "fully adequate" unless it could have "served as a basis for predicting the event in question."[43]

Beyond philosophical issues stemming from the coupling of explanation and prediction in H&O's DN model of explanation, space was cleared for the return of the concept of emergence as faith in reductionism faltered. The DN model of explanation was well suited to reductively explain intratheoretic phenomena but it was soon realized that it was poorly suited to reductively explain intertheoretic phenomena. It makes sense to speak of the deducibility of the explanandum from the explanans if they share similar or identical terms, as would be the case if the phenomena are understood according to the same general theory, i.e., an intratheoretic reduction. It is however quite an odd feature of the DN model that completely different terms may appear in the explanans and explanandum despite the latter being logically deducible from the former. This is the case of intertheoretic reduction whereby a phenomenon is describable in completely different terms by two significantly different theories. Consider the reductive explanation of thermodynamic phenomena in terms of the kinetic theory of matter, or the explanation of chemical phenomena in terms of quantum theory.[44] In short, it seems odd to maintain that thermal laws, governing thermodynamic phenomena for example, are "reduced"

[42] Ibid., 353-354.
[43] Hempel and Oppenheim 1948, 12.
[44] Nagel 2008, 364.

to laws figuring in the general kinetic theory of matter given that terms like 'heat' in thermodynamics are absent from the kinetic theory of matter. Not only do these considerations suggest a decoupling of explanation from prediction in the DN model, but it also raises difficulties surrounding the way in which such explanations ought to be considered reductive. Intertheoretic reductive explanations do not possess a deductive form that intratheoretic reductive explanations possess because of the consistent terminology inherent in the latter. As a proponent of the DN model of reductive explanation, Nagel's proposed solution to remedy this deductive deficit with intertheoretic reductions was to introduce the concept of bridge laws.

> If [intertheoretic] reductions are to be subsumed under the general pattern of scientific explanations [i.e., the DN model of explanation], it is clear that additional assumptions must be introduced as to how the concepts characteristically employed in the reduced laws, but not present in the reducing theory, are connected with the concepts that do occur in the latter.[45]

So in the case of intertheoretic reductions, for example the reductive explanation of physical optics to electromagnetic theory, bridge laws are additions to the explanans that allow for the logical deducibility of the explanandum by connecting distinct terms appearing in each component of the explanation.

As prediction was separated from explanation so too was explanation separated from reduction. Increasing interest in the study of so-called chaotic systems demonstrated that explanation and prediction are not tightly coupled. Immeasurable differences in the initial conditions of a system for example could result in unpredictable macroscopic behaviour. Strict accounts of reduction were similarly seen as inadequate, even fundamentally flawed, to describe

[45] Ibid.

the relatedness of different phenomena. Despite initially appearing as a successful remedy to the difficulties of intertheoretic reduction, Nagelian bridge laws ultimately failed to adequately relate deduction, i.e., reduction, to explanation. The appropriateness and availability of Nagelian bridge laws occupied center stage in debates over reduction and reductionism, with no clear resolution to these problems. The result is that philosophical and scientific space was cleared for the introduction of theories of nonreductive physicalism, "a doctrine that aspires to position itself as a compromise between physicalist reductionism and all out dualisms."[46]

No longer a term only applicable in the realm of epistemology, the concept of emergence has been resurrected in the late 20th and early 21st century in the wake of mounting philosophical unease surrounding reduction and empirical evidence suggesting that the scale and complexity of phenomena can induce qualitative changes in kind, not just degree. Emergent phenomena are increasingly seen as possessing ontological import. This is not only because of problems inherent in the idea of reduction, but also because despite reductionisms successes, they by no means imply success for constructivism. In short, if by reductionism we understand that a phenomenon is nothing more than the fundamental constituents of that phenomenon, constructivism implies the converse: from fundamental constituents, we can construct the phenomenon of interest. The failure of constructivism, more than the failure of reductionism, suggests that ontologically novel and distinct phenomena do exist, and emergence is a concept we can use to understand these ontologically novel phenomena. In contrast to what the early British Emergentists thought, emergent phenomena are not necessarily unexplainable nor are they strictly (and binarily) unpredictable. Moreover, the antithesis of emergence is not reduction; the two concepts are in fact compatible in different ways. Certain phenomena, mind

[46] Kim 1999, 4.

for example, may be ontologically emergent but epistemically reducible. Similarly other phenomena, bird flocks for example, may be epistemically emergent but ontologically reducible. Other aspects of emergent phenomena have been retained as they were introduced by the early British Emergentists. The relative abundance of emergent phenomena as well as a separation of phenomena into "levels" or "domains" are ideas still closely associated with the concept of emergence just as they were at the turn of the 20th century. Although emergence is a difficult concept to articulate and still faces difficult challenges, such as the problem of mental causation, I aim to provide a positive account of ontological emergence in the next chapter of this thesis.

Chapter 2

What Is There, Really?

Scientific advances over the past century have not just produced an abundance of technologies for human consumption, they have also illuminated the universe we inhabit. Our universe is old, roughly 14 billion years old. It is filled with fundamental particles[47] with conservative estimates reaching 1×10^{80} particles. To be clear, that is the number one followed by eighty zeroes. The observable universe[48] is also immense, stretching 93 billion lights years in diameter. It is a

[47] At this point "fundamental particles" can be understood as those particles in the currently accepted standard model of particles physics that cannot be further decomposed into constituent particles. Fundamental particles therefore include quarks, leptons, and gluons. Protons and neutrons, although fundamental for our understanding of the atom, are composed of quarks, and so are not fundamental particles.

[48] Never mind the un-observable universe, which I will not be speculating on. Experimental observations of the expansion of our universe preclude the possibility of our interacting with certain parts of this universe based on their distance from the Earth and given a cosmic speed limit of interaction, i.e., the speed of light. Thus although the observable universe stretches an impressive 93 billion lights years across, the part of the universe accessible to humanity continues

universe in time[49] and ordered according to certain laws of interaction. It is no wonder that after observing the consistent rising and setting of the sun and moon, ordered progression of seasons and march of stars across the night sky that thinkers across time have seen an order in the universe. The Earth is but one tiny island in a vast cosmic ocean.

For the curious student of the 21st century, none of these facts should be surprising. However scientists and philosophers continue to debate over the details of the universe I just painted. At the moment I am sitting at my desk writing this chapter of my MA thesis. My desk has a variety of properties including the property of being coloured and the property of being solid. I perceive my desk to be coloured light brown and I similarly sense its solidity. My elbows do not penetrate the surface of the desk as I rest them on it, nor do my books or laptop fall through or sink into the desk. Although my desk appears to be a single object, there is an important sense in which it is not a single object. This would be especially clear if we were to scale ourselves down to the size of atoms. My desk is composed of billions of objects, atoms, linked in certain ways. Moreover, these smaller objects that constitute the object that is my desk do not possess properties that my desk does. Colour for example, is a property not attributable to single atoms. Solidity similarly becomes more difficult to understand at the atomic level considering that common sense intuitions of solidity find little purchase on atomic facts. For example, if by "solid" we mean "impenetrable," how are we to reconcile this with the fact that atoms are little more than empty space? Specifically, 99.9% of the mass of an atom resides in its nucleus which is often 100,000 times smaller than the size of the atom itself, which can range in

---

to shrink as the universe expands. Current estimates identify this cosmic event horizon as being a mere 32 billion light years in diameter. This is in contrast to the particle horizon, that 93 billion light year swath of universe that could potentially, and indeed does, interact with humanity.

[49] Although I will elaborate on this concept, by the phrase 'a universe in time' I simply mean that this is a universe of change. From the spin flipping of certain fundamental particles to the birth and death of stars, ours is a universe brimming with interaction and change.

size from 32 to 225 picometers (this is six orders of magnitude smaller than the size of a human hair, which can range in size from 17 to 180 micrometers).[50] Sir Arthur Eddington makes this point in his introduction to *The Nature of the Physical World* when he remarks that everyday he sits down at his two tables.

> One of them has been familiar to me from earliest years. It is a commonplace object of that environment which I call the world...It has extension; it is comparatively permanent; it is coloured...Table No. 2 is my scientific table...It does not belong to the world previously mentioned—that world which spontaneously appears around me when I open my eyes...My scientific table is mostly emptiness. Sparsely scattered in that emptiness are numerous electric charges rushing about with great speed; but their combined bulk amounts to less than a billionth of the bulk of the table itself.[51]

How are we to understand the relationship between these two tables? Are there two tables that exist or only one? Is one “more real” than the other? Does the existence of one determine or depend on the existence of the other or are they mutually determinative of a single object? Can we reduce one to the other, thereby eliminating one to insist only on the existence of the other? Equally, if we start with one of the tables, can we construct the other out of the “stuff” of the one we start with?

Metaphysical Preliminaries

Eddington’s two tables bring into sharp relief the differences between how the world appears to us and how it actually is, at least as understood through the modern scientific method. Although certain properties of objects appear to us to persist and exist as if they were independent of

[50] The Physics of the Universe, “The Universe By Numbers.”
[51] Eddington 1968, xi-xii.

other objects or properties, this is often not the case. Colour, for example, like the colour of the blue folder on my desk, is dependent, among other things, on the interaction between light and the molecular surface of my folder. I perceive the colour blue when looking at my folder because it reflects electromagnetic (EM) radiation of a wavelength that causes my perception of blue[52] and absorbs other wavelengths of EM radiation. This being the case, although it appears to me that the "blueness" of my folder is independent of the object that is my folder, this is not true. Rather the colour is dependent at least on the properties of the molecular surface of my folder. Note that the relationship between the object denoted by the term 'folder' and its property of being blue is even independent of my folder being a folder. That is to say, I am able to cut my folder in half or otherwise modify it such that it can no longer perform the function I require of it, and throughout all of these changes, so long as the properties of its molecular surface remain unchanged,[53] it will still be coloured blue.

Returning to Eddington's two tables, and given the account of colour sketched above, we are in a position to begin to understand the relationship between the table of everyday experience and the scientific table. Just as colour is determined by the molecular surface of an object, so too are many of the properties of the table of everyday experience determined by the properties of the scientific table. In addition to colour, the property of solidity that my desk possesses is determined by the properties of the connections between the molecules that constitute the object that is my desk. So in response to the question, "Does the existence of one table determine or depend on the existence of the other?" we may answer in the affirmative.

---

[52] The perception of the colour blue corresponds to EM radiation possessing a wavelength approximately between 380nm and 500nm.

[53] Although I say "unchanged" this is not entirely true. The properties of the molecular surface may change insofar as it retains the property of primarily reflecting EM radiation between 380nm and 500nm while absorbing all other visible wavelengths of EM radiation.

The colour of the table of everyday experience often changes when scratched since such an action significantly alters the molecular surface of the table so that different wavelengths of EM radiation are reflected and absorbed, resulting in change in perception of the colour of the table. Similarly if we were to attempt to loosen the connections between the molecules that constitute the object that is my desk, by heating it up for example, we would witness the table of everyday experience lose the property of solidity as it changes state from solid to liquid, or perhaps even a gas. These considerations suffice to show that we may also answer affirmatively to the question of whether one table may be reducible to the other. The table of everyday experience appears, to use a reductionist slogan, to be "nothing more than" the scientific table. In other words, any properties of or changes to the object that is the table of everyday experience are wholly determined by the properties of or changes to the scientific table.

This view of properties and objects presupposes a monistic metaphysical commitment to physicalism, the barest version of which simply asserts that all that exists is physical. There are no entelechies, Cartesian minds or immaterial substances. Indeed the committed physicalist need not be troubled by the accounts of colour and solidity given above, and may even happily eliminate colour and solidity from our ontology, relegate them to the status of a mere epiphenomenon or, more charitably, consider them as ontologically secondary. However there are other properties that are not as easily reduced by virtue of a dependency relation to properties that are ontologically prior. Moreover, it is not clear whether, even in the presence of a dependence relation, a given property ought to be considered for elimination from our ontology or given some other status. Strict versions of reductionism, when dealing with relations of dependence, often entail either eliminativism, i.e., an elimination of a given property or property type from our ontology, or else entail identity, i.e., an identification of a given property or property type with a more fundamental property or property type. Again, we may be satisfied by this view of the world in regards to properties such as being coloured or being solid, however there are other properties, such as being ferromagnetic or being superconductive, that resist reduction either via elimination or identity that motivate a different view of properties in the world.

Emergence

The popularity of reductionism stems from its association with other attractive metaphysical commitments such as ontological minimalism and constructivism. Humphreys explains the connection between these ideas.

> Ontological minimalism runs roughly like this: a) There is a relatively small set of fundamental constituents of the world; b) to individuate these we need only intrinsic (i.e., non-relational) properties; and c) all the non-fundamental individuals and properties are composed of or from these fundamental entities. Ontological minimalism is an attractive view. Indeed, if one accepts a well-known set of philosophical doctrines, it comes out almost right. Chief amongst these doctrines are 1) giving primacy to logical reconstructions of ordinary and scientific concepts, and 2) a broadly Humean account of causal relations.[54]

However, as discussed at the end of Chapter 1, there is reason to doubt the tenability of the constructivist project. Reductionism qua reductionism has certainly been successful. The discoveries of quantum mechanics and molecular/genetic biology arose partly out of efforts to understand how phenomena could be understood on the basis of more fundamental components. These reductive triumphs however were also seen as vindicating constructivism. It

[54] Humphreys 1997, S337-S338.

was argued that the genome is all that was needed, for example, to entirely reconstruct a living organism. Although this is true in a certain sense, mounting empirical evidence demonstrating the difficulty of reconstructing complex phenomena from fundamental constituents illuminated a growing gap between reduction and construction. So despite the success of reductionism, reductionism itself does not entail constructivism. In short, despite our ability to reduce everything to fundamental physical constituents, i.e., fundamental physical objects, properties, causes, laws, types or kinds, it does not follow that from those fundamental physical constituents we can reconstruct the universe and all of the phenomena within it. We are now in an era of emergence because the "twin difficulties of scale and complexity" have stymied the constructivist project and brought our attention to the fact that the behaviour of large and complex aggregates of fundamental constituents "is not to be understood in terms of simple extrapolation of the properties of a few particles."[55] Rather, at differing levels of complexity wholly new phenomena appear (emerge!) whose understanding requires study as fundamental as any other. This emergentist view however does not come free of charge. One must be willing to reject both ontological minimalism and constructivism[56] and instead embrace ontological pluralism.

The goal of this chapter is therefore to articulate a defensible account of the concept of emergence. Specifically, I will be arguing for an account of ontological emergence. In short, there exist objects and properties (phenomena) in addition to laws that are not understandable purely on the basis of fundamental constituents and the fundamental laws governing those constituents. Emergent phenomena constitute a novel ontological domain governed by domain-

[55] Anderson 1972, 393.

[56] More specifically, one must reject a strong version of constructivism, i.e., the view that all things can be reconstructed from fundamental constituents. Weaker versions of constructivism may be compatible with the concept of emergence.

specific laws that are neither eliminable nor identifiable, via reduction, with other ontological domains. Importantly, this is not to say that there are no phenomena that can be reduced to and successfully reconstructed from fundamental constituents and laws. The weakest versions of the concept of emergence only entail rejecting ontological minimalism and embracing ontological pluralism. Essentially, there exists both emergent and non-emergent phenomena.

The account of ontological emergence advanced here is based on the account given by Paul Humphreys outlined in his book *Emergence: A Philosophical Account*, albeit with certain significant changes. The descriptive account of ontologically emergent phenomena that will be maintained consists of four characteristics: relationality, novelty, irreducibility and broken symmetry (or synonymously, spontaneously broken symmetry). An often overlooked aspect of emergent phenomena is that they "result from something else," that is, emergent phenomena are fundamentally relational in nature.[57] Emergent phenomena also "possess a certain kind of novelty with respect to the [phenomena] from which they [emerge]."[58] Ontologically emergent phenomena are irreducible (ontologically), via elimination or identity, to the phenomena from which they emerge. Lastly ontologically emergent phenomena exhibit broken symmetry. The normative claim that will be maintained in this thesis is that the first characteristic, relationality, is necessary in all cases of emergence (e.g. necessary for ontological emergence and epistemological emergence) whereas the remaining three characteristics, novelty, irreducibility and broken symmetry, are necessary in cases of ontological but not epistemological emergence, although they may be present in cases of epistemological emergence. Let us examine each of these characteristics in turn.

[57] Humphreys 2016, 26.
[58] Ibid.

Emergence: Relationality

All approaches to understanding emergence conceive of emergent phenomena as relational phenomena. A seemingly trivial point, "the fact that emergent entities result from something else" is of crucial importance.[59] Emergent phenomena are phenomena that are necessarily related to other phenomena. Although not explicitly stated in the writing of the early British Emergentists, this characteristic of emergent phenomena is nevertheless present. C. Lloyd Morgan's investigation of consciousness for example draws attention to the relationship between the body of an organism and its consciousness.

> ...if we assume that some of the vital processes of the animal organism are correlated with conscious experience, we have to face the questions: If some, why not all? We have to consider the problem of the relation of life to consciousness throughout the whole range of organic evolution and development...Or it may be that consciousness appeared later than life, and if so we have to face the questions: When, from what source, and under what conditions?[60]

Elsewhere Morgan is explicit about the metaphysical assumptions, specifically a commitment to monism, he is working under. The relational aspect of emergent phenomena figures prominently in his discussion and is framed in terms of the emergence of consciousness, or mind, from the world given that the existence of mind was the paradigmatic example of an emergent phenomenon at the time.

> This disparity [between the world we are conscious of and consciousness itself] is, it will be urged, fundamental. The problem of philosophy is to explain how these two utterly

---

[59] Ibid., 27.
[60] Morgan 1912, 89.

> diverse existences come into relationship—not the relationship of part to whole within one order of existence; nay, rather of *this* mind-order with *that* world-order. But the assumption on which I proceed is that there is, for scientific treatment, *one* order and only one. Within that order there are many and varied relationships—and among these relationships are those which we call experiential or conscious.[61]

What needs highlighting here is the fact that emergent phenomena are necessarily and inextricably linked to other phenomena. If consciousness is emergent, as many British Emergentists thought, or if non-functional aspects of consciousness, such as qualia, are considered to be emergent, then they are so by virtue of their relationship to the body or specific parts of the body. But there are many kinds of relations that can hold between domains of things, such as between physical properties and mental properties. Psychophysical parallelism for example is the doctrine that the physical domain is separate from the mental domain but that events in the two domains occur in a synchronized fashion. Emergence is not an intrinsic feature of a phenomenon, so under the assumptions of psychophysical parallelism mind would not be considered an emergent phenomenon because it has no relation to the physical domain.

The account of ontological emergence I am advancing relies on a robust sense of relationality. In short, "the primary relata in the ontological approach are the emergent [phenomenon] and the [phenomena] from which it emerges," allowing for property emergence, object emergence and nomological emergence.[62] The property of being superconductive for example emerges from the property of being able to occupy the same quantum state.[63] Likewise biological objects emerge from abiotic objects and "higher-level" laws emerge from "lower-

[61] Ibid., 133.
[62] Humphreys 2016, 42.
[63] A detailed examination of the ontological emergence of the property of superconductivity appears at the end of this chapter.

level" laws.[64] Relationality however is also necessary for epistemological emergence. In contrast to ontological emergence, by epistemological emergence we understand our knowledge of a property, object or law. The primary relata in cases of epistemological emergence are therefore not phenomena, i.e., properties, objects or laws themselves, but rather concepts denoting phenomena, i.e., concepts denoting properties, objects or laws. As a result we must be careful to distinguish between a phenomenon itself and the concepts through which we come to have knowledge of a phenomenon. Moreover, it is possible to maintain the epistemic emergence of a phenomenon whilst simultaneously maintaining that the phenomenon is ontologically non-emergent. Consider the example of joint probability distributions. Knowledge of the joint probability distribution for two random variables is not determined by the distributions of the variables taken separately.

> Taking the two examples together [a system of two fair coins that are probabilistically independent and a system of two coins welded at a point on their circumference so that they always come up with the same side], this shows that the probabilities of the individual events alone do not determine the probabilities of the joint events. As a result, the joint distribution does not supervene on the marginal distributions alone...the joint distribution does not exist without a specification of the interactions between the systems that generate the individual probabilities.[65]

It is not clear that this example of the joint probability distributions is evidence for ontological emergence.[66] Although the joint probability distribution "emerges" in some sense from the systems that generate the individual probabilities, this seems to be a case of epistemic

[64] Talk of "levels" is common when discussing concepts like emergence, reduction and the relatedness of different ontologies. However whenever possible I will be eschewing talk of levels in favour of "domains."
[65] Humphreys 2016, 73-74.
[66] This is primarily because the novelty criterion necessary for ontological emergence is not met, however this will be elaborated upon in the next section.

emergence as opposed to ontological emergence. Notice for example that the fact that there exists a joint probability distribution does not entail our knowledge of the joint probability distribution. Any inferences we might make about the joint probability distribution, specifically, knowledge we have about the joint probability distribution, is inextricably linked to our knowledge of the composite welded coin system. Thus the joint probability distribution is epistemically emergent in part because of our inability to make predictions about the system from knowledge of the constituent systems alone.[67] Importantly, contra the early British Emergentists, from this it does not follow that we cannot explain this phenomenon. This is a key difference between the account of emergence I am advancing and the account articulated by the early British Emergentists. Indeed an unfortunately erroneous connection between the idea of unexplainability and the concept of emergence, one that I aim to dispel, persists to this day.[68] Instead of signaling the unexplainability of a phenomenon, the epistemic emergence discussed above merely denotes that "knowledge about the emergent [phenomenon] [in this case the joint probability distribution] must be obtained from sources other than R [the theory or model used to represent the system under consideration]."[69] In this case, the other sources from which we can obtain knowledge about the emergent joint probability distribution are likely a combination of observation and experimentation on other similar systems.

[67] Humphreys 2016, 38.
[68] See Chapter 4 for more details on the connection between emergence and explanation.
[69] Humphreys 2016, 38.

Emergence: Novelty

Emergent phenomena are in an important sense "novel" phenomena. Novelty as a characteristic of emergent phenomena was explicitly noted by early British Emergentists such as Morgan.

> We live in a world in which there seems to be an orderly sequence of events. It is the business of science, and of philosophy which keeps in touch with science, to describe the course of events in this or that instance of their occurrence, and to discover the plan on which they proceed. Evolution, in the broad sense of the word, is the name we give to the comprehensive plan of sequence in all natural events. But the orderly sequence, historically viewed, appears to present, from time to time, something genuinely new...Salient examples are afforded in the advent of life, in the advent of mind, and in the advent of reflective thought.[70]

The radical novelty of certain phenomena was a primary motivator for different thinkers, including the early British Emergentists, to conceive of the concept of emergence. The chemical properties of water for example differ in a number of significant ways from the chemical properties of its constituent elements hydrogen and oxygen. So ontologically emergent phenomena are relatively, not absolutely, novel insofar as the emergent phenomena is compared to phenomena from which it emerges. This idea of relative novelty is brought out in the writing of Samuel Alexander, another early British Emergentist.

> The argument is that mind has certain specific characters to which there is or even can be no neural counterpart. It is not enough to say that there is no mechanical counterpart, for the neural structure is not mechanical but physiological and has life. Mind is, according to our interpretation of the facts, an 'emergent' from life, and life an emergent from a lower physico-chemical level of existence. It may well be that, as some

[70] Morgan 1923, 1.

> think, life itself implies some independent entity and is indeed only mind in a lower form.[71]

Yet despite the recognition that a necessary characteristic of emergent phenomena is the appearance of something new, how we ought to understand this "newness" is still debated.

Over a century after the British Emergentists attempted to articulate the idea that emergent phenomena are novel in this or another sense, scientists and philosophers continue to debate just how this idea of novelty ought to be cashed out. Positions range from Paul Teller's outright rejection of novelty as a useful concept (since "novelty proves unhelpful as soon as we press for more precision")[72] to Alexander Rueger's precise stipulation of novelty couched in the language of dynamical systems theory (DST) (a property is novel if, when comparing two systems, the behaviour of the non-reference system in which the property of interest manifests is itself qualitatively different from or topologically inequivalent to the behaviour of the reference system).[73] Some philosophers, like Jaegwon Kim, attempt to rehabilitate the notion of novelty employed by the early British Emergentists.

> I believe that "new" as used by the emergentists has two dimensions: an emergent property is new because it is unpredictable, and this is its epistemological sense; and, second, it has a metaphysical sense, namely that an emergent property brings with it new causal powers, powers that did not exist before its emergence.[74]

Although it is tempting to connect novelty to the idea of unpredictability, as Kim does, novelty and unpredictability are separate and distinct ideas that must remain so, not least because

[71] Alexander 1920, 14.
[72] Teller 1992, 140.
[73] Rueger 2000, 303.
[74] Kim 1999, 8.

novelty is necessary for ontological but not epistemological emergence. A detailed investigation of the concept of "novelty" will likely involve an analysis of other closely related concepts such as the observer, system under study, as well as an accounting of quantitative measures of complexity, all of which is simply beyond the scope of this paper. I raise these examples of definitions of novelty for two reasons: i) to insist on novelty as a necessary characteristic of emergent phenomena, and ii) to demonstrate that there is a general consensus affirming the possibility of ontologically novel phenomena.

So if there are theoretical and empirical reasons to accept novelty as a legitimate necessary characteristic of emergent phenomena, how ought we understand the concept of "novelty"? Humphreys offers a definition that I believe is both precise and general enough for our purposes.

> An entity E is novel with respect to a domain D just in case E is not included in the closure of D under the closure criteria C that are appropriate for D. Novelty can be taxonomic novelty in the sense that the entity is not included in a given conceptual classification.[75]

This definition is precise because it simultaneously emphasizes the necessary characteristic of relationality, i.e., a phenomenon is novel "with respect to" other phenomena, while also protecting against a relativistic infinite regress of emergent phenomena vis-a-vis the criterion of novelty (without the closure clause, identification of novelty would lead to an infinite regress of observers detecting novelty, detecting novelty, etc). Moreover, its generality allows it to encompass most, if not all, of the articulations of the concept of novelty commonly associated with the concept of emergence. Consider Kim's reinterpretation of the early British Emergentists

[75] Humphreys 2016, 29.

position, “that an emergent property brings with it new causal powers, powers that did not exist before its emergence.”[76] Kim is specifically interested in the emergent phenomena of mind and the causal powers mental phenomena are purported to have. This being the case, we can see how, according to the definition of novelty given above, the causal powers of mind might be considered novel. Mental causation (the entity E) is not included in the closure of the physical domain (D) under closure criteria (C) appropriate for the physical domain. Although in this particular example I might be accused of begging the question (mental causation is assumed to be novel despite that being the thing that needs establishing), what I hope to have shown is that there is a sufficiently adequate definition of novelty which can be applied to potentially ontologically emergent phenomena.

Emergence: Irreducibility

The third necessary characteristic of ontologically emergent phenomena is their irreducibility to the phenomena from which they emerge. Importantly, from the ontological irreducibility of a phenomenon it does not follow that the same phenomenon is also epistemically irreducible. Just as emergence can be understood ontological or epistemically, so too can reducibility, i.e., reduction, be understood ontologically or epistemically.[77] Moreover, one and the same phenomenon can be ontologically reducible and epistemically irreducible or vice versa. However the phenomena we are most interested in are emergent phenomena which are ontologically irreducible and can be either reducible or irreducible epistemically.

[76] Kim 1999, 8.
[77] See Chapter 1 for a discussion on the relation between ontological and epistemological reduction.

As mentioned in the previous chapter, the basic idea of irreducibility, insofar as we are interested in its use in the context of emergence, can be conveyed by the "something more than…" slogan. Recall that early British Emergentists argued that mind requires something more than merely neural processes. The basic idea of reducibility can conversely be conveyed by the "nothing more than…" slogan. The mass of an object for example is nothing more than the sum of the masses of its constituents. What these slogans obscure however are two fundamental ambiguities concerning the nature of the reductive relation. The first ambiguity concerns the relata of the relation, specifically what types of things are linked by a reductive relation. Ontologically, as was the case with the relata of emergent relations, reductive relations concern properties, objects and laws. The property of being coloured blue for example can be reduced to the property of reflecting EM radiation of a specific wavelength. Likewise aggregated objects (e.g., a pile of sand) are reducible to their constituents and context specific laws are reducible to more general laws (e.g., the reduction of Galileo's or Kepler's law to Newton's laws of motion). Epistemically, as was the case with emergence, reductive relations concern not the phenomena themselves but our concepts of the phenomena. The primary relata in cases of epistemic reductions include our concepts of properties, objects and laws, but can also include larger epistemological constructs such as theories, models or representational frameworks.[78]

The second and more important ambiguity, since it determines whether we categorize a particular phenomenon as emergent or not, concerns how the ontological relata in a reduction relation are linked. Four major ontological links have been proposed: elimination, identity, composition and supervenience.[79] Each will be examined in turn and all will ultimately be

[78] Van Gulick 2001, 4.
[79] Ibid.

rejected as inadequate to characterize the kind of ontological relation emergent phenomena and the phenomena from which they emerge share, hence the ontological irreducibility of emergent phenomena. Elimination is perhaps the most straightforward kind of reductive relation "in which we come to recognize that what we thought were Xs are really just Ys."[80] For example, while there was a time during human history when solar eclipses were considered to be the acts of a vengeful deity, we now know such events are simply the result of the moon casting a shadow on the Earth as it passes between the Earth and sun. Vengeful sun-eating deities can therefore be reduced, via elimination from our ontology, to celestial mechanics. Emergent phenomena, such as the emergence of a superfluid from liquid helium, exhibit the opposite kind of relation. That is, the ontological relation invoked in cases of emergent phenomena is precisely a relation that denies that Xs are really just Ys. Emergence is invoked when Xs are manifestly not Ys.

In contrast to elimination, reduction via identity involves "cases in which we continue to accept the existence of Xs but come to see that they are identical with Ys (or with special sorts of Ys)."[81] Colour is an example of a phenomenon that can, ontologically, be reductively identified with EM radiation of a certain wavelength (or range of wavelengths). The phenomenon of heat can similarly be reductively identified with mean kinetic molecular energy, but we nevertheless maintain the existence of heat just as we maintain the existence of colour. Emergent phenomena however are not merely existent, rather emergent phenomena possess a novelty not possessed by the phenomena from which they emerge. The ontological emergence relation is antithetical to reduction via an identity relation. Although colour is an ontologically non-

[80] Ibid.
[81] Ibid., 5.

emergent phenomenon when it is considered in relation to EM radiation, we can understand colour to be emergent, i.e., non-identical, when considering its relation to individual atoms. The property of being coloured is not a property we can attribute to single atoms, only to sufficiently large aggregates of atoms, hence the emergence of colour from atoms rather than the identification with and subsequent reduction of colour to atoms. Heat is emergent in just the same way given that it is a property attributable only to a collection of molecules and not individual molecules. The existence of colour and the existence of heat is not identical with the existence of an atom and a molecule.

Reduction however may also proceed via appeal to the composition of a phenomenon and thereby avoid issues associated with identity. In essence, if all of the parts of a phenomenon are of a certain kind, then it appears there is justification to maintain that that phenomenon is nothing but a phenomenon of the same kind (e.g., if all of the parts of mind are neurophysiological, then mind is nothing more than the neurophysiological). However as Van Gulick notes, in the context of mind-body dualism, "to say that a thing is composed *entirely of physical parts* is not the same as saying that *all its parts are entirely physical*," which is needed in order to exclude the possibility of nonphysical phenomena, either at the level of the parts or the level of the whole.[82] Composition is therefore insufficient to ground a reductively physicalist worldview given its compatibility with many different versions of dualism, such as psychophysical parallelism. The emergence relation in contrast is both compatible with a commitment to physicalism (admittedly not an ontologically minimal version of physicalism) and incompatible with the different versions of dualism. This is because emergence relations require the appropriate kind of precedence between the emergent phenomenon and the phenomena

[82] Ibid., 7.

from which it emerges. For example a distinctive characteristic of emergent relations are their irreflexivity, i.e., emergent phenomena are of a fundamentally different kind in comparison to the phenomena from which they emerge.[83] Organismal fitness for example emerges from the organism-environment interaction and is distinctively irreflexive in the sense that organismal fitness, as an existent phenomenon, is neither an element of the organism itself nor is it an element of the organism's environment.

The ontological relation between two phenomena may also proceed according to a supervenience relation. Supervenience has been, and continues to be, an appealing way to make sense of ontological reduction. Accounts of the supervenience relation usually link sets of properties thereby avoiding problems associated with reduction via composition whilst simultaneously retaining those aspects of the composition relation that make it more attractive than the identity relation. Specifically supervenience attempts to reductively link sets of properties without appealing to either identity or composition such that "one set of properties (X-properties) supervenes on another (Y-properties)."[84] Reduction is achieved in the sense that supervenience requires a strict asymmetrical dependence between the supervening and subvenient properties such that there is an invariant correlation between the two. In this way the appearance of the supervening property is nothing more than an appearance of its determining subvenient base. However similar to composition, the supervenience relation is too weak to ground reductive physicalism given its compatibility with dualistic positions like Descartes' interactionist dualism or Spinoza's panpsychist monism. Supervenience applies in any case in which two sets of properties covary in a determined way "whether [this is] underwritten

[83] Humphreys 2016, 28.
[84] Van Gulick 2001, 7.

by natural law (*nomic supervenience*) or some stronger metaphysical link (*metaphysical supervenience*).”[85] Furthermore, despite the similarities between the supervenience relation and the emergence relation, such as their characterization via the presence of an asymmetry, emergence does not entail supervenience nor does supervenience entail emergence.[86] One reason for this is the fact that emergent phenomena are intrinsically temporal whereas supervening phenomena are atemporal. We say that solid ice emerges from liquid water, not that it supervenes on liquid water, in part because of the temporal evolution necessary to transition from one phase of matter to another. On the other had we say that solidity supervenes on a crystal lattice, i.e., the solid phase of matter, because there is no temporal component mediating the appearance of solidity given the formation of a crystal lattice. In short, at the same instance that a crystal lattice forms it can be said to possess the property of being solid. Thus a distinctive characteristic of the emergence relation that separates it from the supervenience relation is that it is fundamentally diachronic whereas supervenience is typically construed as either an atemporal or synchronic relation.

Relations of elimination, identity, composition or supervenience simply do not capture the relation between the relata in cases of emergence. Fundamentally physical phenomena only provide a necessary condition for the existence of emergent phenomena. In other words, the mere existence of fundamentally physical phenomena does not guarantee the existence of phenomena that emerge from the fundamentally physical, hence a rejection of the supervenience relation. Even if that were the case, i.e., even if the existence of fundamentally physical phenomena guaranteed the eventual existence of mind for example, it does not follow

[85] Ibid., 8.
[86] The differences between emergence and supervenience will be elaborated on in Chapter 3 when challenges facing the concept of emergence are addressed.

that mind is therefore fundamentally physical, contra the composition relation. Indeed the emergence relation's raison d'etre is that the composition relation appears unfruitful and unable to support potentially emergent phenomena (recall emergence is invoked when a phenomenon appears to be something "more than" the phenomenon it is directly related to). Finally, relation via elimination or identity is in direct conflict with the notion that the two relata in an emergence relation are equally existent. Neither the emergent phenomenon itself nor the phenomenon from which it emerges possess a privileged ontological status; neither are eliminable nor are they identical. Ontologically emergent phenomena are necessarily irreducible to the phenomena from which they emerge.

Emergence: Broken Symmetries

Thus far I have avoided (or attempted to avoid) talk of phenomena as stratified or layered or as being part of some natural hierarchy. A common view that comes from the sciences is that the world is composed of levels or is layered in a hierarchical fashion. Particle or fundamental physics is seen as the base, the layer above that being chemistry, above which is biology, psychology and social psychology. Indeed this hierarchy can be compressed or expanded as one desires. We could equally conceive of the world as constituted by elementary particles/properties, atoms, molecules, cells, multicellular organisms, and social organisms. It could be tempting to think that an investigation of broken symmetries (in the context of emergent phenomena) should conceive of a world in a hierarchical fashion, but I will eschew this terminology in favour of the use of the term 'domain': when moving from one domain to another domain we may encounter broken symmetry. In the context of emergence, use of the term ‘domain’ has significant advantages over use of terms like ‘level’ or ‘layer.’ For one, by speaking (or writing) in terms of domains we can sidestep the issue of needing to specify and justify the ontologically fundamental level(s). In a similar way we also avoid issues related to justifying how or why one phenomenon, or a type of phenomena (i.e., property type or object type), is ontologically superior or inferior in comparison to another. This is of critical importance within the context of emergence since neither of the relata in an emergence relation possesses a unique ontic status. Biochemical phenomena are neither ontologically inferior or superior to quantum mechanical phenomena. Indeed emergent phenomena, although historically associated with changes in “levels” of phenomena (e.g., mind “emerges” from life since mind is on a different, higher, ontological “level” than life), can exist both between domains as well as within a single domain. Life for example is a case of inter-domain ontological emergence, between the abiotic and biotic domains, whereas superconductivity is a case of intra-domain ontological emergence, i.e., the emergence of superconductivity is entirely within the domain of fundamental particle physics.[87] Thus another advantage of using the term ‘domain’ is that it is conceptually free of the baggage associated with the term ‘level.’ The connection then between broken symmetry, emergent phenomena, and different ontological domains is this: broken symmetry is a necessary characteristic of emergent phenomena. But this broken symmetry can occur within or between ontological domains corresponding to the potential emergence of a phenomenon within or between ontological domains.

Why broken symmetry? Recall that the concept of emergence is not merely juxtaposed with the concept of reduction but also with the concept of construction (or, synonymously, reconstruction). A rejection of constructivism follows from the emergentist’s presupposed

[87] Superconductivity will be looked at in much greater detail later in this chapter.

rejection of ontological minimalism. In the same way that irreducibility is a necessary characteristic of ontologically emergent phenomena that distinguishes such phenomena from ontologically reducible phenomena, broken symmetry is a necessary characteristic of ontologically emergent phenomena that distinguishes such phenomena from ontologically constructible phenomena. As mentioned at the end of Chapter 1, acceptance of the existence of ontologically emergent phenomena has been largely driven by a recognition that as a system increases in scale and complexity, differences in degree can spontaneously become differences in kind. In short, "a shift from quantitative to qualitative differentiation takes place" hence the idea of broken symmetry "may be of help in making more generally clear the breakdown of the constructivist converse of reductionism."[88] Let us consider this in the context of an example from the field of biology.

The proper functioning of enzymes, proteins that catalyze biochemical reactions, depends crucially on their tertiary and quaternary structures, and these structures are determined by the primary structure of a protein. This primary structure is simply the sequence of amino acids[89] in a protein and represents one level of complexity. This primary structure however determines the secondary structure[90] of a protein as the amino acids in the primary structure begin to interact with one another. Once the secondary structure has formed, the

[88] Anderson 1972, 393.

[89] Amino acids, like sugars, fats and nucleic acids, are a group of organic compounds that have an amine and carboxyl functional group in addition to a variable (R group) side chain specific to each of the 20 unique amino acids. Amino acids are linked via their amine and carboxyl groups to form peptide chains which constitute the primary structure of a protein. That is, the primary structure of a protein is simply the order in which amino acids are linked to form a polypeptide chain.

[90] The secondary structure of a protein refers to the highly regular local sub-structure of the polypeptide chain. These structures take two main forms: the alpha-helix or the beta-strand/beta-sheet and are defined by patterns of hydrogen bonding (interactions between hydrogen and other negatively charged elements, usually oxygen or nitrogen) that arise as a result of the primary structure, i.e., the sequences of amino acids in the polypeptide chain.

protein begins to exhibit an even greater degree of complexity as the fully three dimensional tertiary structure[91] forms. Beyond the tertiary structure, some proteins only function as a subunit in a protein complex that represents the highest level of protein complexity: the quaternary structure.[92] Hemoglobin, the protein in red blood cells responsible for shuttling oxygen throughout the body, is an example of a protein with a quaternary structure since it is composed of four protein subunits[93] that together allow it to fulfill its function.[94] What this protein example serves to illustrate is that reconstruction of phenomena from fundamental constituents can become increasingly difficult, to the point of becoming impossible, as a result of the increasing complexity of a system. In this protein example, by taking single amino acids as fundamental constituents of a protein, reconstruction of the tertiary or quaternary structure becomes increasingly more difficult as the length and variation of amino acids in the primary structure increases.[95] Moreover, this example highlights how a shift from quantitative to qualitative differentiation, and hence symmetry breaking, takes place. Whereas the primary structure of a protein is differentiated on the (quantitative) basis of the sequence of amino

---

[91] Although already “three dimensional,” the tertiary structure of a protein refers to the globular shape that forms as a result of hydrophobic moieties among the secondary structures interacting driving non-specific folding of the protein. The final folded protein is stabilized by strong molecular bonds specific to the tertiary level such as hydrogen bonds, disulfide bonds and salt bridges.

[92] Individual proteins that have already folded into their tertiary structures can interact further with other proteins to form larger protein aggregates which constitutes the quaternary structure of a protein.

[93] The most common hemoglobin type is a tetramer called hemoglobin A since it has four protein subunits, two alpha and two beta subunits, bonded together.

[94] I warn the reader that this explanation of protein structure formation is oversimplified.

[95] Over six million unique protein, i.e., amino acid, sequences have been discovered, and this number continues to rise. Despite this, only a fraction of these sequences have had their corresponding tertiary structures experimentally determined, at just over 50,000. This is primarily because computational methods grounded in determining tertiary structure using only the amino acid sequence itself are either computationally intractable or else fail to adequately predict protein structure (except for those with very short sequences). In fact, research on protein structure has been increasing outsourced to the public via the creation of protein folding “games” in which users are allowed to manipulate proteins in search of protein homologies/analogies that better indicate (in contrast to amino acid sequence) the final protein structure. See, Kelley and Sternberg 2009, for more information.

acids, the tertiary structure (and quaternary structure) of a protein is differentiated in a wholly different manner, namely on the (qualitative) basis of its function in a biochemical process.

Instead of examining the complexity of protein structure, Philip Anderson notes this same characteristic of symmetry breaking through the increasing complexity, and subsequent qualitative change in behaviour, of chemical compounds such as ammonia ions as they dynamically invert via quantum mechanical tunneling.

> The symmetry involved in the case of ammonia is parity, the equivalence of left and right-handed ways of looking at things...Hydrogen phosphide, PH3, which is twice as heavy as ammonia, inverts, but at one-tenth the ammonia frequency. Phosphorus trifluoride, PF3, in which the much heavier fluorine is substituted for hydrogen, is not observed to invert at a measurable rate, although theoretically one can be sure that a state prepared in one orientation would invert in a reasonable time. We may then go on to more complicated molecules, such as sugar, with about 40 atoms. For these it no longer makes any sense to expect the molecule to invert itself. Every sugar molecule made by a living organism is spiral in the same sense, and they never invert, either by quantum mechanical tunneling or even under thermal agitation at normal temperatures. At this point we must forget about the possibility of inversion and ignore the parity symmetry: the symmetry laws have been, not repealed, but broken.[96]

We can understand symmetry as “the existence of different viewpoints from which [a] system appears the same.”[97] Thus both the protein example and Anderson’s own ammonia example serve to highlight how increases in the complexity of a system, i.e., increases in the amount of “stuff,” namely amino acids and atoms respectively, can lead to broken symmetries. In the case of protein structure, the symmetry that is broken is not parity but rather functionality. The primary structure of a protein, the mere sequence of amino acids, is non-functional. Yet once

[96] Anderson 1972, 394.
[97] Ibid.

the tertiary (or quaternary) structure has formed the protein becomes functional[98] and may carry out its biochemical purpose within an organism. Despite the fact that the primary structure of a protein does not change after the tertiary structure has formed, and despite the fact that changes in protein structure are entirely within the same ontological domain, symmetry has nevertheless been broken. This is because the system, the protein, no longer performs the same function when viewed from its tertiary structure as opposed to its primary structure: it appears significantly different in regards to its functionality.

Just as there are a variety of ontologically emergent phenomena (e.g., superconductivity, protein functionality, mind, first order phase transitions, ferromagnetism, etc), so too are there a variety of broken symmetries that can accompany the appearance of ontologically emergent phenomena. The first order phase transition from liquid to crystalline solid, which is the type of transition that takes place when liquid water freezes to form solid ice, involves symmetry breaking (as does the reverse transition, from solid ice to liquid water). Specifically, continuous translational and rotational symmetry are broken in the transition from liquid to crystalline solid.[99] In contrast to liquids, crystalline solids can sustain or resist a shear stress (they are rigid) and are only symmetrical under discrete translations or rotations, a result of "the degree to which the molecules are no longer arbitrarily oriented [as they are in a liquid or gas] but have now oriented themselves along a specific direction."[100] Recall that, in general, if a system appears the same when viewed from two different perspectives, we call that system symmetrical insofar as its sameness is preserved when switching between the two perspectives.

[98] I warn the reader once again that this is greatly oversimplified. Some proteins, even after folding into their tertiary or quaternary structures, require "activation" by other proteins to become fully bioactive.
[99] Anderson 1981, 4-6.
[100] Ibid., 6.

So we ascribe continuous translational symmetry to a system if we are able to translate (i.e., move through space) a view of that system, by any amount, and still observe the same view of the system. Consider a canister of liquid or gas as an example system. In that canister the molecules are arranged arbitrarily and interacting randomly such that if we were to take a particular view of the system, perhaps at one end of the canister, such a view would be indistinguishable from any other translated view of the system. Continuous rotational symmetry (indeed all symmetries) arises in just the same way. We ascribe continuous rotational symmetry to a system when we are able to rotate a view of that system, by any amount, and still observe an identical view of the system. A system that transitions from a liquid phase to a solid phase (e.g., the freezing of liquid water) violates continuous translational and rotational symmetry. This is because solid phases of matter, in contrast to liquid and gaseous phases, exhibit regular, non-arbitrary and non-random, spatial organization such that if we translate or rotate a view of the system in a certain way, the view of the system will no longer be identical with the previous perspective. Importantly, there are certain specific translations or rotations that can be made to systems that are solids (specifically crystalline solids) so that our view of the system remains the same after making such translations or rotations. This is why crystalline solids do not violate translational or rotational symmetry wholesale, rather they only violate the continuous aspect of the symmetry. We ascribe discrete translational or rotational symmetry to a system if we are able to translate or rotate a view of that system, by specific discrete amounts, and still observe the same view of the system. The presence of any kind of broken symmetry is a necessary characteristic of ontologically emergent phenomena. Solid ice is therefore ontologically emergent with respect to the liquid water from which it emerges (assuming it has also met the other three necessary characteristics of emergent phenomena outlined above) because continuous translational and rotational symmetry has been broken. Indeed a strength of this account of emergence that separates it from Humphreys' is that symmetry breaking is modellable within dynamical systems. As Anderson notes, "matter will undergo mathematically sharp, singular "phase transitions" to states in which the microscopic symmetries, and even the microscopic equations of motion, are in a sense violated."[101]

In fact, Anderson speculates that there may be many different forms of symmetry breaking when moving between different ontological domains (although instead of 'domain' he uses terms such as 'stage' and 'hierarchy').

> It seems to me that the next stage [in unsymmetric phenomena] is to consider the system which is regular but contains information. That is, it is regular in space in some sense so that it can be "read out," but it contains elements which can be varied from one "cell" to the next. An obvious example is DNA; in everyday life, a line of type or a movie film have the same structure. This type of "information-bearing crystallinity seems to be essential to life...one further phenomenon seems to be identifiable and either universal or remarkably common, namely, ordering (regularity or periodicity) in the time dimension...temporal regularity is a means of handling information, similar to information-bearing spatial regularity...In some sense, structure—functional structure in a teleological sense, as opposed to a mere crystalline shape—must also be considered a stage, possibly intermediate between crystallinity and information strings, in the hierarchy of broken symmetries.[102]

The relevance and significance of the parallels between Anderson's proposed broken symmetries and emergent phenomena, such as the emergence of protein functionality, supports the view that the presence of a broken symmetry (or symmetries) in a system is necessary to maintain the emergent status of a phenomenon in that system.

---

[101] Anderson 1972, 395.
[102] Anderson 1972, 395-396.

A Case Study: The Emergence of Superconductivity

Ontologically emergent phenomena are characterized by four necessary features: relationality, novelty, irreducibility and broken symmetry. Epistemologically emergent phenomena require only the first feature, relationality, although many examples of epistemologically emergent phenomena also exhibit some combination of the remaining three features. (e.g., novelty, irreducibility, or broken symmetry). In this section I will closely examine the phenomenon of superconductivity and argue that, since it satisfies the four necessary features outlined above, it ought to be considered a genuine case of ontological emergence.

Superconductivity, before it was observed, was a property neither predicted nor imagined to exist given that the very idea of a conductor with zero resistivity seemed to violate the known laws of physics. In non-superconducting metals such as copper, electrical resistance drops as the temperature of the conductor is lowered. Even as the temperature nears absolute zero in non-superconducting materials, electrical resistance never becomes zero. This is a result of a fundamental physical law known as the Pauli Exclusion Principle which governs the behaviour of electrons (the particles that carry an electric current). Superconductors however do not behave in this way. In contrast to non-superconducting materials, as the temperature of a superconductor is lowered it eventually reaches a critical temperature at which the Pauli Exclusion Principle is violated resulting in the pairing of electrons whose behaviour mimics that of an altogether different kind of fundamental particle known as a boson.[103] Bosons do not obey the Pauli Exclusion Principle and as a consequence there is no limit on the number of bosons that may occupy the same quantum mechanical state. The resultant quantum mechanical

[103] Schon 1990, 239.

coherent superconducting electron pair[104] is responsible for many different macroscopic effects such as the Meissner Effect, Josephson Effect, and superconductivity, the latter being characterized by the property of having zero electrical resistance, or more accurately, the property of being able to sustain a persistent electrical current for any measurable[105] time.[106]

First, superconductivity exhibits the relationality characteristic of ontologically emergent phenomena. The relata we are interested in are the emergent phenomenon itself, superconductivity in this case, and the phenomena from which it emerges, namely a specific kind of system. Only systems configured using certain materials, such as mercury or aluminum, and subjected to certain conditions, such as being cooled below its critical temperature (4.1 Kelvin for mercury or 1.2 Kelvin for aluminum), are capable of supporting the property of superconductivity. The novelty characteristic of ontologically emergent phenomena is also met in the case of superconductivity. The property of superconductivity is novel with respect to the domain of classical electromagnetism and is not included under the closure criteria appropriate for classical electromagnetism, one of which is the criterion that electrical resistivity can never drop to zero. The irreducibility characteristic of ontologically emergent phenomena is similarly satisfied. The properties of the paired electrons that give rise to superconductivity are irreducible to the properties of a single electron, or indeed any property of leptons[107] in general,

[104] Schon 1990, 242.
[105] Although the electrical current in a superconductor will eventually decay, this time is calculated to be orders of magnitude larger than the predicted lifespan of the universe.
[106] This is a simplified explanation of the phenomenon of superconductivity. For an in-depth exploration of the quantum mechanical phenomenon and the competing theories that purport to explain it see both, Anderson 1959 and Hirsch 2009.
[107] According to the currently accepted standard model of particle physics, there are two main categories of particles: bosons, which mediate the transmission of fundamental forces, and fermions, which are the fundamental particles themselves. Fermions are further subdivided into quarks, the particles that constitute atomic nuclei, and leptons, of which the electron is the most common member.

all of which must obey the Pauli Exclusion Principle. This also highlights another perspective from which the property of superconductivity is novel: with respect to the domain of electron behaviour,[108] violation of the Pauli Exclusion Principle is prohibited under the closure criteria appropriate for electron behaviour. Lastly, superconductivity exhibits the necessary characteristic of broken symmetry, specifically spontaneously broken gauge symmetry.[109] In short, the transition in a system from a non-superconducting state to a superconducting state, usually achieved by lowering the temperature of the system below its critical temperature, precipitates a phase change corresponding to a spontaneously broken symmetry.[110] In the case of superconductivity, previously unpaired electrons obeying the Pauli Exclusion Principle violate the principle, i.e., break gauge symmetry, by pairing together (known technically as Cooper pairs) in the transition to a superconducting state. The paired electrons no longer behave as electrons and instead behave as if they are a boson, a different kind of fundamental particle that is subject to different fundamental laws.[111] In sum, given the above considerations, we can insist

[108] The behaviour of electrons and all fundamental particles with half-integer spin (an intrinsic property of all fundamental particles classified as fermions, including quarks and leptons (e.g., electrons)) are governed in part by the spin-statistic theorem of quantum mechanics which relates spin to the particle statistics a fermion, such as an electron, obeys.
[109] Leggett and Fernando 1991.
[110] This is a simplified view of broken symmetry. One may be wondering how exactly symmetry is measured, and the answer is against a complex object (i.e., a “view” of a system that is normally invariant given certain changes to the system captured by a specific complex number, vector or tensor) known as the order parameter. In general the order parameter captures the degree of organization of a system. That organization however can change drastically if a system reaches a critical point and undergoes a discontinuous phase change such that a previously apparently conserved quantity is no longer conserved. For a more technical understanding of symmetries see, Brown and Holland 2004.
[111] Bosons, in contrast to leptons, do not obey the Pauli Exclusion Principle and as a result there is no restriction on the number of bosons that may occupy the same quantum state. This characteristic of bosons is not just responsible for the property of superconductivity (where pairs of electrons behave as if they are a boson) but also the property of superfluidity, another candidate ontologically emergent phenomenon that possesses the property of having zero viscosity.

on the genuine ontological emergence of the property of superconductivity. In the next chapter I will seek to defend the concept of emergence from its most serious challenges.

Chapter 3

Emergence and Mind

The attractiveness of the thesis that some phenomena are emergent when it comes to understanding their ontic status stems from its generality. It is applicable to the quantum mechanical phenomena of superconductivity, superfluidity and ferromagnetism and to the complex macroscopic phenomena of phase transitions in matter, life and the phenomenon of mind. Despite this, the concept of emergence is beset by the challenges of causal exclusion and downward causation, though they are ultimately surmountable. Recall that the account of ontological emergence being defended here is under the monistic metaphysical assumption of physicalism. In regard to the specific phenomenon of mind and the potential emergence of mind from physico-biochemical phenomena, the account advanced here is one version of what Mario Bunge calls "psychoneural monism," specifically "emergentist [physicalism]."[112] Indeed most criticisms of the concept of emergence are centered around the phenomenon of mind and whether other kinds of relations are better suited to describe the relationship between mental phenomena and physical phenomena, broadly speaking. One of the more attractive alternatives

[112] Bunge 1977, 504-509.

to emergence is supervenience.[113] In short, the concept of supervenience involves the dependence of one set of properties on another set of properties. For example, we can think of the property of roughness (or smoothness) as supervening on the properties of the molecular surface of an object. Colour is another property that can similarly be seen as supervening on the properties of the molecular surface of an object. In contrast to emergence, supervenience is a kind of reductive relation. That is, the supervening property is reducible to the subvenient property, i.e., the property or properties that determine the supervening property.

Critics of emergence insist that the emergence relation inadequately characterizes the relation between the physical and the mental, and this is for two main reasons. First, the novelty of emergent phenomena, but specifically the novel causal powers an emergent phenomenon may have, appear to compete with the causal powers of the emergent base. If mental phenomena are indeed emergent with respect to physico-biochemical phenomena (prima facie, mental phenomena satisfy the conditions of relationality, novelty, irreducibility and broken symmetry), then the emergentist must contend with the challenge of mental causation. Jaegwon Kim explains:

> The problem [of mental causation] arises from what I call “the supervenience argument.” This, I claim, is our principal problem of mental causation. In referring to this as “our” problem of mental causation, what I mean to suggest is that it is a problem that arises for anyone with the kind of broadly physicalist outlook that many philosophers, including myself, find compelling or, at least, plausible and attractive...The fundamental problem of mental causation for us, then, is to answer this question: How is it possible for the mind to exercise its causal powers in a world that is fundamentally physical?[114]

[113] See Chapter 2 for a detailed discussion of the connection between reduction, supervenience and emergence.
[114] Kim 1998, 30.

There are many reasons to insist that mental phenomena do have causal powers. Human agency for example, is in serious danger of elimination from our ontology or relegation to the status of an epiphenomenon if it is the case that "our beliefs, desires, and intentions" do not have causal effects in the world.[115] Under strict reductionist schemes, just as solidity or roughness become difficult to reconcile with atomic facts, so too would human agency become difficult to understand according to the physico-biochemical facts of human neurophysiology. In short, if it were not the case that "in voluntary actions our beliefs and desires, or intentions and decisions...somehow cause our limbs to move in appropriate ways, thereby causing objects around us to be rearranged," we would have no reason to insist on the reality and ontological novelty of mental causation.[116] In addition to human agency, Kim notes that the possibility of human knowledge is also tied to the reality of mental causation. If it is the case that our perceptions, reasoning processes, and memory are involved in the way in which we acquire new knowledge and beliefs from our existing reservoir of what we already do know and believe, and if it is also the case that "memory is a complex causal process involving the interactions between experience, their physical storage, and retrieval in the form of belief," the following must be the case.[117] Taking away the mental phenomena of perception, reasoning, and memory also takes away the possibility of all human knowledge, except for perhaps instinctive or reflexive knowledge, if those can even be considered "knowledge."[118]

---

[115] Ibid., 31.
[116] Ibid.
[117] Ibid.
[118] We might not consider instinctive or reflexive knowledge as "knowledge" because they fail to satisfy the classic justified true belief (JTB) account of knowledge insofar as the justification and/or belief criteria may not be met. When reflexively pulling one's hand away from a hot stove for example, it is unclear whether a belief that "the stove is hot" has been formed before the reflex has occurred.

Problems with Emergence: Mental Causation and The Exclusion Problem

Mental causation is a specific example of two more general problems that threaten to undermine both the usefulness and coherence of the concept of emergence. This section and the subsequent section will address the connections between mental causation and the general problems of causal exclusion and downward causation, and their connection to emergence. In the third and fourth sections I will defend the concept of emergence and argue that the causal exclusion and downward causation problems can be overcome.

The first general problem is the causal exclusion problem, and it follows from three general claims:

> (1) If an event *x* is causally sufficient for an event *y*, then no event *x** distinct from *x* is causally relevant to *y* (*exclusion*).
> (2) For every physical event *y*, some physical event *x* is causally sufficient for *y* (*physical determinism*).
> (3) For every physical event *x* and mental event *x**, *x* is distinct from *x** (*dualism*).
> (4) So: for every physical event *y*, no mental event *x** is causally relevant to *y* (*epiphenomenalism*).[119]

In the specific case of mental phenomena, the causal exclusion problem manifests as an explicit rejection of the causal efficacy of mental phenomena. The issue is considerably more serious than that however. In the generalized form of this argument the problem becomes reconciling how the causal powers of any supervening property are related to the causal powers of its subvening property. The generalized causal exclusion problem can be recovered by substituting "mental event *x**" for "supervening event *x**" in claims (3) and (4) from the above argument,

[119] Yablo 1992, 247-248.

which has the potential to threaten the causal efficacy of all events (e.g., chemical, biological, geological, etc) that supervene on the fundamentally physical subvenient base (hereafter, "causal exclusion" will refer to the argument just outlined).

The coherence of the concept of emergence is threatened by the causal exclusion argument primarily because emergentists appear to endorse the claims in the causal exclusion argument, claims which conflict with the concept of emergence (in the sense that the emergentist should reject claims (1) and (3)). Explicitly, by insisting on the ontic status of emergent phenomena, of which the emergence of mind from a physico-biochemical base is a paradigmatic example, the emergentist appears to be endorsing the dualism claim, that mental phenomena are distinct from physical phenomena. Implicitly, monistic accounts of emergent phenomena appear to endorse both the exclusion and physical determinism claims. In short, the emergentist appears, prima facie, to be committed to mind-body supervenience in the sense that mental phenomena are distinct from and causally dependent on physical phenomena. Indeed this is precisely the line of reasoning Kim uses to argue that the emergentist must contend with the problem of causal exclusion. Kim maintains that the generalized causal exclusion problem stems from the concept of supervenience itself, and he motivates this conclusion by examining the supervenience of the mental on the physical.

> Suppose then that mental event *m*, occurring at time *t*, causes physical event *p*, and let us suppose that this causal relation holds in virtue of the fact that *m* is an event of mental kind *M* and *p* an event of physical kind *P*. Does *p* also have a physical cause at *t*, an event of some physical kind *N*? To acknowledge mental event *m* (occurring at *t*) as a cause of physical event *p* but deny that *p* has a physical cause at *t* would be a clear violation of the causal closure of the physical domain, a relapse into Cartesian interactionist dualism which mixes physical and nonphysical events in a single causal chain. But to acknowledge that *p* also has a physical cause, *p**, at *t* is to invite the question: Given that *p* has a physical cause *p**, what causal work is left for *m* to

> contribute? The physical cause therefore threatens to exclude, and preempt, the mental cause. This is the problem of causal exclusion.[120]

If Kim is right, then the challenge facing the nonreductive physicalist, and consequently the emergentist, is to give an account of how two supposedly sufficient causes of one and the same event are related to each other. To acknowledge that there are two sufficient causes for a particular mental phenomenon, for example, is to confront the problem of causal overdetermination. Conversely to maintain that there exists only one sufficient cause for a particular mental phenomenon is to either endorse full blown idealism on the one hand, i.e., all that exists is mind, or reductive physicalism on the other, i.e., all that exists is physical.

Kim maintains that mind-body supervenience is the "minimal commitment that anyone who calls herself a physicalist should be willing to accept," and since the account of emergence being defended here is within a physicalist framework, our focus shall be on understanding the similarities and differences between the concepts of supervenience and emergence.[121] To begin, Kim presumably maintains that mind-body supervenience is the minimal commitment a physicalist must accept given that a rejection of mind-body supervenience is tantamount to a rejection of physicalism itself, at least insofar as the relationship between mind and matter is concerned. Accepting that the mental does supervene on the physical, then at best we can regard mental phenomena as epiphenomena bereft of causal efficacy as per the causal exclusion argument given above (i.e., from the three claims of exclusion, physical determinism and mind-matter dualism, it follows that all mental phenomena are epiphenomena). To reject the supervenience of the mental on the physical is, at least according to Kim, to relapse into more

---

[120] Kim 1998, 37.
[121] Ibid., 38.

robust ontological dualisms, such as Cartesian interactionist dualism or psychophysical parallelism, both of which would be serious missteps for a proposed physicalist theory of mind. Accepting then that the relation between mind and matter is one of supervenience, we may define supervenience in the following way.

> Mental properties *supervene* on physical properties, in that necessarily, for any mental property *M*, if anything has *M* at time *t*, there exists a physical base (or subvenient) property *P* such that it has *P* at *t*, and necessarily anything that has *P* at a time has *M* at that time.[122]

More generally, supervenience between two sets of properties A and B, not just mental and physical properties, can be similarly defined.

> A...*supervenes* on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and *necessarily* if any y has G, it has F.[123]

In short, supervenience is a pattern of covariance connecting two sets of properties such that identical subvenient bases correspond to identical supervening states. Further, supervenience implies an asymmetric synchronic dependency or determination component between the relata such that the supervening state is dependent upon the subvenient base, or equally, that the subvenient base determines the supervening state.[124] The dependence operates in a single direction given that the subvening property's appearance does not depend on the supervening property's appearance. Further, although the relation between property sets that share a

[122] Kim 1998, 9.
[123] Kim 1993, 65.
[124] Kim 1998, 11.

supervenience relation is usually thought of atemporally, specific instances of supervenience, i.e., mind-body supervenience, are conceived synchronically. In other words, the moment a subvenient property appears is the same moment the supervening property appears.

> For example, if a person experiences pain, it must be the case that that person instantiates some physical property (presumably, a complex neural property) such that whenever anyone instantiates the physical property, she must experience pain. That is, every mental property has a physical base that guarantees its instantiation. Moreover, without such a physical base, a mental property cannot be instantiated.[125]

In Kim's example, the mental phenomenon of pain is the higher level property that supervenes on the lower level property, albeit a "complex neural" one. The example clearly highlights both the determination and covariance components of the supervenience relation. All mental properties are "guaranteed," i.e., determined by, a physical base. Moreover, the determination is asymmetrical given that without the complex subvenient neural property the corresponding mental property of pain would not arise. Similarly the determination is synchronic since necessarily anything that has that complex neural property "at a time" has the corresponding supervening mental property of pain "at that time." Finally, any changes to the subvenient physical neural property results in covarying changes to the supervening mental property. The famous case of Phineas Gage, an American railroad construction worker, is evidence for the pattern of covariance between the physical and mental: after an iron rod was driven through his head destroying most of his left frontal lobe, anecdotal evidence suggests Gage's personality was significantly altered following the accident.[126]

---

[125] Ibid., 9-10.
[126] Gage is but one example in which a physical change to the brain has been thought to be responsible for significant changes to personality traits. Following a horrific killing spree, an

By extension, just as the causal powers of supervening mental properties are threatened by the causal powers of the subvening physical properties, Kim maintains that the causal powers of emergent phenomena, such as mind, are threatened by the causal powers of the phenomena from which mind emerges, such as its neurophysiological base. To be sure, emergence is similar to supervenience in the sense that the emergence relation also has an asymmetric dependence component. The emergence of superconductivity requires an emergent base that can support the appearance of such a property, but this dependence is not a synchronic one. Emergent phenomena are necessarily diachronic phenomena, e.g., if an emergent base appears at t, then the phenomenon that emerges from that base appears at $t+n$.[127] Furthermore, contra Kim, it is not clear that emergence entails supervenience nor is it clear that we would be worse off rejecting mind-body supervenience which results, at least according to Kim, in a rejection of the causal closure of the physical domain and an acceptance of mind-body dualism. These issues will be taken up again in the section titled *In Defense of Emergence*.

Problems with Emergence: Mental Causation and Downward Causation

The second general problem that threatens the concept of emergence is the problem of downward causation. The specific example of mental causation can also be subsumed under the

autopsy performed on Charles Whitman discovered a tumour in his brain that, conceivably, could have contributed to his inability to control his emotions and actions. Given the lack of understanding of the relationship between neurophysiological changes and personality/behaviour /mental changes that persists to this day, these conclusions are speculative at best. However what is certain is the fact that there is a relationship between neurophysiology and mental phenomena.

[127] By the variable n I mean to represent the epistemically significant unit of time through which we understand the system, and consequently the emergent phenomenon, under scrutiny. For example, if we are interested in the emergent phenotypic changes to a population of organisms then n may represent a single generation.

general problem of downward causation which emergentists and all nonreductive physicalists must address. As in the case of causal exclusion, Kim draws attention to the fact that emergentism implies downward causation, and he does so via the specific example of mental causation.

> But why are emergentism and nonreductive physicalism committed to downward causation, causation from the mental to the physical? Here is a brief argument that shows why. At this point we know that, on emergentism, mental properties must have novel causal powers. Now, these powers must manifest themselves by causing either physical properties or other mental properties. If the former, that already is downward causation. Assume then that mental property M causes another mental property M*. I shall show that this is possible only if M causes some physical property. Notice first that M* is an emergent; this means that M* is instantiated on a given occasion only because a certain physical property P*, its emergence base, is instantiated on that occasion. In view of M*'s emergent dependence on P*, then, what are we to think of its causal dependence on M? I believe that these two claims concerning why M* is present on this occasion must be reconciled, and that the only viable way of accomplishing it is to suppose that M caused M* by causing its emergence base P*. In general, the principle involved here is this: *the only way to cause an emergent property to be instantiated is by causing its emergence base property to be instantiated*. And this means that the "same-level" causation of an emergent property presupposes the downward causation of its emergent base. That briefly is why emergentism is committed to downward causation. I believe that this argument remains plausible when emergence is replaced by physical realization at appropriate places.[128]

On its own, the problem of downward causation implies that emergent phenomena are causally efficacious only insofar as they cause the next (causally speaking) emergent base to appear. Yet such a narrow conception of ontologically emergent phenomena conflicts with both the intuition behind and formal characterization of ontological emergence advanced in the previous chapter. Intuitively, ontologically emergent phenomena are "more than," in a nontrivial sense, the phenomena from which they emerge. If emergentism requires downward causation, then that

[128] Kim 1992, 136.

"something more" can at most, as per the downward causation argument, be construed as an intermediate causal step linking fundamentally physical emergent bases. This seems altogether weaker than the sense in which ontologically emergent phenomena are thought of as "more than" their emergent bases. Indeed if the concept of emergence entails downward causation, then the necessary criterion of irreducibility for ontologically emergent phenomena appears in danger of collapsing into a reducible relation of identity. In Kim's example above, if mental property M causes mental property M*, but only does so by causing physical property P*, which is M*'s emergent base, why insist on M's causal powers at all? This especially in light of the fact that M itself is instantiated by its physical emergent base P. If emergent properties supervene on their emergent bases in this way, as Kim suggests they must, then it follows from the problem of downward causation that the irreducibility of emergent phenomena to their emergent bases is untenable.

Couple the problem of downward causation to the problem of causal exclusion and emergentists will find themselves facing a serious contradiction. It follows from the causal exclusion argument that there is no mental event causally relevant to a physical event, and from the downward causation argument it follows that emergentism implies a causal connection between mental and physical events. The problem of mental causation exemplifies both of these more general problems. On the one hand, if mind, as an emergent phenomenon, supervenes on the physical, its emergent base, and both are causally efficacious, we must deal with the issue of assigning causal precedence, i.e., we must confront the problem of causal exclusion. On the other hand, given that emergentism as a doctrine reifies emergent phenomena, such as the mind, mental phenomena necessarily have causal powers. These causal powers however must operate in a "downward" direction, i.e., mental phenomena do not cause the appearance of other mental phenomena on the same “level,” rather they indirectly cause mental phenomena to appear by directly causing the corresponding physical emergent base, hence the problem of downward causation.

In Defense of Emergence

Despite the above considerations, I maintain that the causal exclusion and downward causation problems can be addressed using two different strategies. First, if it is the case that the causal exclusion problem stems from the very concept of supervenience, as Kim argues it does, then if we can demonstrate that the concept of emergence does not entail supervenience, we can establish that the concept of emergence does not entail the exclusion problem (vis-a-vis supervenience at any rate). Second, both the causal exclusion and downward causation problems can be dissolved by giving up a commitment to the causal closure of the physical domain. Importantly, the causal closure of the physical domain is distinct and separate from the concept of the causal completeness of the physical domain. It is only a rejection of the causal completeness of the physical domain, not a rejection of the causal closure of the physical domain, that would result in a relapse into psychophysical dualisms such as Cartesian interactionist dualism.

In Defense of Emergence: Emergence Does Not Imply Supervenience

Although emergence and supervenience are related,[129] emergence does not entail supervenience in general or mind-body supervenience in particular. A necessary characteristic of emergent phenomena is that there is an irreducible relation between the emergent phenomenon and the base from which that phenomenon emerges. This relationship however is not necessarily one of strict dependence or determination[130] as is the case with supervenience. The existence of physico-biochemical phenomena for example does not automatically imply the existence of the emergent phenomenon of mind, for example, that emerges from physico-biochemical phenomena. Moreover, there is a tension between the determination and covariation components in the supervenience relation that is not present in the emergence relation. Specifically, the strict synchronic dependence of the supervening property on the subvenient base is in danger of collapsing into a relation of identity, especially if we insist on the reality of the supervenient properties. Humphreys illustrates this using a concrete example.

> The property approach [to understanding supervenience] introduces a presumption that the properties F and G are real and that G is not identical to F...It is appropriate to consider a circle as a degenerate case of an ellipse, with eccentricity zero. To do so facilitates a neat classification and allows an orderly taxonomy within the theory of conic sections. In contrast, a definition of the divisor operation that does not exclude division

[129] See Chapter 2, specifically the section titled *Emergence: Irreducibility*, for a discussion of the relation between emergence and supervenience within the context of reducibility.

[130] A strict dependence, such as logical or metaphysical necessity, is not required for ontological or epistemological emergence which require only a much looser nomological necessity. In contrast supervenience, especially the more common version of supervenience called "strong supervenience," is typically conceived of as implying at least a metaphysical or logical/conceptual necessity. Granted there are weaker versions of supervenience (aptly named "weak supervenience") which may settle for nomological necessity.

> by zero is defective, not the least reason being that there is a discontinuous change in properties of the division operation when the second argument is zero.[131]

In other words, there are some examples of supervening properties in which the supervenience relation collapses into an identity relation. This is the case for the supervenient property of eccentricity, specifically the eccentricity of a circle wherein the property of "having exactly zero eccentricity" just is the property of "being a circle." If we insist on the reality of both properties (of "having eccentricity equal to zero" and of "being a circle") then they are related not via supervenience but via identity. The same cannot be said for the supervening property of "being a divisor" which exhibits the relation typical of a supervenience relation. The property of "being a divisor" depends asymmetrically, but more importantly synchronically, on the subvenient property of "being a real number," with the important caveat that that real number cannot be zero. The supervenience relation between the two relata, being a divisor and being a real number, does not degenerate into an identity relation. The worry that the dependence between an emergent phenomenon and the base from which it emerges may collapse into an identity relation is simply not present, as it is in the case of supervenience. For one, this is because the emergence relation is a fundamentally diachronic relation. That is, if we really wanted, we could individuate an emergent phenomenon and its emergent base on the basis of temporal evolution. More important however is the fact that emergent phenomena are not reducible phenomena. They are reducible neither via a relation of supervenience nor are they reducible via a relation of identity. This is because the concept of emergence is invoked precisely in situations in which an irreducibly novel phenomenon has arisen, i.e., an emergent phenomenon is one that appears not to be identical to the phenomenon from which it emerges.

---

[131] Humphreys 2016, 213.

Beyond the fact that the supervenience relation itself is, in some scenarios, in danger of collapsing into a relation of identity, emergence just does not entail supervenience in general nor does it entail mind-body supervenience in particular. Two examples should together suffice to demonstrate why this is the case. Recall that from mind-body supervenience it follows "that necessarily any two things (in the same or different possible worlds) indiscernible in all physical properties are indiscernible in mental respects."[132] In short, mind-body supervenience, aimed as it is at relating an inferior "level" of phenomena to a superior "level" of phenomena, turns into claims about how inferior parts are related to superior parts of a single whole. We say that many, if not all, properties of a whole supervene on the properties of the parts of that whole since "the properties of the whole [the mental, neurophysiological, biochemical, and physical properties for example] are determined by the qualitative intrinsic properties of the most fundamental parts" of that whole.[133] Evidence of phenomena in which the properties of the whole are not determined by the intrinsic properties of the most fundamental parts will therefore demonstrate the failure of supervenience and consequently the failure of mind-body supervenience. Indeed there is such evidence, with the first example coming from the realm of quantum mechanics.

> It frequently has been noted that one of the distinctive features of quantum states is the inclusion of non-separable states for compound systems, the feature that Schrodinger called "quantum entanglements." That is, the composite system can be in a pure state when the component systems are not, and the state of one component cannot be completely specified without reference to the state of the other component. Furthermore, the state of the compound system determines the states of the constituents, but not vice versa. This last fact is exactly the reverse of what

---

[132] Kim 1998, 10.
[133] Silberstein 2001, 68.

> supervenience requires, which is that the states of the constituents of the system determine the state of the compound...[134]

Certain quantum mechanical systems appear to resist description via supervenience in regard to the relationship between the system considered as a whole and the constituents of that system. These systems exhibit the irreducibility characteristic of ontologically emergent phenomena. Moreover, these systems appear to exhibit downward causation given that "the state of the compound system determines the state of the constituents, but not vice versa." Since ontological emergence entails both an irreducible relation and downward causation, and supervenience entails the opposite, i.e., both a reducible relation and epiphenomenalism, it follows that emergence does not entail supervenience in general. In short, the phrases "emergent property M emerges from the emergent base P" and "supervening property M supervenes on subvenient base P" are not equivalent. But what of the specific example of the relation between mind and body? Mind has been offered both as an example of an ontologically emergent phenomenon, emerging from neurophysiology, and as an example of the supervenience of one set of properties, mental properties, on another set of properties, perhaps complex neural properties. But are there any examples or evidence for the failure of supervenience insofar as mind and it subvenient physical base are concerned? I believe there are, which leads us to the second example which demonstrates that emergence of mind from the body does not entail the supervenience of mind on the body.

In arguing for the causal efficacy of mental phenomena, Silberstein draws on a particular abnormal psychology phenomenon outlined in the DSM IV (Diagnostic and Statistical Manual IV)

[134] Humphreys 1997a, 15-16.

called "Conversion Disorder."[135] Briefly, Conversion Disorder is a diagnostic category used in some psychiatric classification systems that is applied to patients who present with neurological symptoms that do not appear to have a physiological cause. Diagnostic features of Conversion Disorder important for our purposes are as follows.

> The symptoms cannot be fully explained by either a neurological/medical condition or by external causes such as substance abuse or environmental/cultural forces; diagnostic testing shows no physical cause for the dysfunction. Conversion symptoms typically do not conform to known anatomical pathways and physiological mechanisms, but rather follow the patient's conceptualization of a condition. The more medically naive the person, the more implausible are the presenting symptoms...Conversion symptoms are often inconsistent. A 'paralysed' extremity will be moved inadvertently while dressing or when attention is directed elsewhere...[136]

As Silberstein notes, "contra crude [reductive] physicalism, the fact that Conversion symptoms come with neurochemical mechanisms does in no way negate the primarily *psychological* aetiology of this disorder."[137] Moreover, contra supervenience, subvenient neurophysiological, i.e., physical, properties appear to be divorced from supervening psychological, i.e., mental, properties. In short, in cases of Conversion Disorder, supervening mental property M appears to have been caused in the absence of its subvenient complex neurophysiological property P. The fact that Conversion Disorder symptoms are not observed to map onto known anatomical pathways coupled with the fact that the symptoms follow the patient's conception, i.e., mental representations, of a condition, suggests that at the very least subvening neurophysiological properties underdetermine corresponding mental properties.

---

[135] Silberstein 2001, 65.
[136] Ibid., 65-66.
[137] Ibid., 66.

If correct, this conclusion neutralizes the problem of causal exclusion and supports the view that mental phenomena are more than mere epiphenomena, they have causal efficacy. Specifically, this conclusion undermines the first and second claims of the causal exclusion argument. Claim (1), exclusion, is undermined because a physical event alone would be causally insufficient to account for either the physical or mental symptoms associated with Conversion Disorder. Claim (2) of the causal exclusion argument, physical determinism, must be explicitly rejected if it is the case that, as diagnostic testing suggests, no physical cause for the dysfunction brought on by Conversion Disorder exists. Admittedly, medical and neurophysiological sciences are far from their ideal limits, and Silberstein acknowledges just that assuming that "as neuroscience grows in technical and diagnostic sophistication it will find underlying neurochemical causal mechanisms or correlates for Conversion symptoms,"[138] as was and still is the case with other mental phenomenon such as Clinical Depression.[139] Nevertheless I maintain, as Silberstein does, that Conversion Disorder offers proof of the causal efficacy of mental phenomena and directly refutes the problem of causal exclusion. Although it is true that supervening properties are causally inefficacious, this is not the case for ontologically emergent phenomena. Empirical evidence suggests that the conclusion of the causal exclusion argument, that no mental event is causally relevant to a physical event, is false, and this is because mental events are not supervenient: they are emergent with respect to underlying neurophysiological events. Mental phenomena, as emergent phenomena, are causally connected to both other mental phenomena and neurophysiological phenomena in a causal chain that spans different

[138] Ibid.

[139] I offer Clinical Depression just as an example of a mental phenomenon that was once miserably understood and had its mysterious aura dispelled by advances in neuroscience and neurophysiology, although this may be a contentious claim. As much as we do know about Clinical Depression, there is still much that we do not know.

ontological domains. The concept of emergence is therefore able to preserve the causal efficacy of an emergent phenomenon (e.g., the causal efficacy of the mind) and thereby avoid the consequences of the causal exclusion argument, which suggests emergent phenomena cannot possess causal powers.

In Defense of Emergence: Emergence Entails A Plurality of Ontologies

So the emergentist can avoid the problem of causal exclusion because the concept of emergence does not entail the concept of supervenience, as Kim thinks it does. Moreover, contrary to what Kim believes, to acknowledge that a mental event is a cause of a physical event is not to relapse into Cartesian interactionist dualism. Accepting the existence of ontologically emergent phenomena presupposes a rejection of ontological minimalism, the idea that there is a fundamental ontological "level" that all phenomena can be reduced to. Indeed the problem of downward causation is no problem at all for the emergentist since the doctrine of emergentism presupposes no privileged ontological domain, and neither does it prohibit causal interaction between different ontological domains. Rejecting the privileged status of certain ontological levels, such as the fundamentally physical level, allows the emergentist to sidestep the problems of both causal exclusion and downward causation as well as make better sense of complex causal connections between phenomena.

Kim maintains that the concept of emergence is flawed because it implies downward causation which in turn implies the interjection of a mental, i.e., nonphysical, cause into a supposedly fundamentally physical causal chain. The conclusion that emergence is flawed however only follows because Kim conflates the causal closure of the physical domain with the causal completeness of the physical domain. In short, Kim argues that to maintain the causal efficacy of an emergent phenomenon one must violate the causal closure of the physical domain, something which we should be loath to do considering our physicalist commitments. Indeed Kim believes that rejecting the causal closure of the physical domain is tantamount to rejecting physicalism itself and amounts to embracing some kind of ontological dualism in regards to the relation between an emergent phenomenon and its emergent base. The picture however is considerably more complex than Kim appears to think.

First, consider that the causal closure of the physical domain is a significantly different and distinct idea from the causal completeness of the physical domain. Humphreys for example defines causal closure as:

> Two events are causally connected just in case one is a cause of the other. Then, a domain D is *causally closed* just in case anything causally connected to an element of D is itself an element of D.[140]

It is true then, at least according to Humphreys' definition of causal closure, that a mental phenomenon able to causally influence a physical phenomenon would be an example of a violation of the causal closure of the physical domain. After all, mental phenomena and physical phenomena are wholly distinct and radically different according to the supervenient view of mind Kim espouses. Indeed claims (2) and (3), physical determinism and dualism respectively, in the causal exclusion argument preclude the possibility of rejecting the causal closure of the physical domain. The concept of emergence however does not entail a commitment to ontological dualism, such as mind-body dualism. On the contrary, the concept of emergence

[140] Humphreys 2016, 31.

presumes a rejection of ontological minimalism in favour of an acceptance of ontological pluralism. The problem of downward causation for the emergentist can therefore be dissolved in two different ways: by recognizing that there is evidence, in ordinary language and scientific explanations, that support the view that causes do in fact regularly leave and reenter the physical domain, and by recognizing that outright rejecting the causal closure of the physical domain, as long as we maintain its causal completeness, is not nearly as threatening as Kim would have us believe. It is tempting to think, as most proponents of supervenience do, and as Kim does, that the concept of emergence implies dualism especially with regard to mental versus physical phenomena, but this is incorrect. The concept of emergence is perfectly intelligible under a monist framework, specifically physicalism.[141] Thus the causal exclusion and downward causation problems can be dissolved by recognizing that the concept of emergence does not entail a rejection of the causal completeness of the physical domain.

Despite the fact that emergence implies the rejection of the causal closure of the physical domain, this does not also imply that emergence is a concept that entails a commitment to Cartesian interactionist dualism, or other ontological dualisms. To see that this is the case, we must distinguish causal closure from causal completeness. Humphreys outlines a definition for the causal completeness of the physical domain that is similar to his definition for causal closure.

> *Definition* A domain is *causally complete* just in case every event in D has a causally sufficient antecedent that is in D.
> This definition allows that causal chains can enter and leave a domain, rendering the domain causally open, even though every event in D can be traced back via a causal chain to some event that is also in D.[142]

[141] Bunge 1977, 508-509.
[142] Humphreys 2016, 32.

Ontological emergence, positioned as it is between radical monism and radical pluralism, is incompatible with the causal closure of the physical domain. It is however compatible with the causal completeness of the physical domain. That is to say, emergent phenomena possess causal efficacy that arises from, but is ontologically irreducible to, their emergent base. Bunge succinctly writes that "what holds for physical systems [e.g., causal powers, property instances, etc.] holds *a fortiori* for chemical, biochemical, biological and social systems."[143] Just as a crystalline solid emerges from a liquid and "enzymatic catalysis is an emergent property" of protein tertiary/quaternary structure, mental phenomena emerge from neurophysiological phenomena.[144] So it is common, in fact all but necessary,[145] for the emergentist to trace causal chains from one domain to another and back again. Ordinary language for example allows that causal chains leave and reenter the physical domain thereby highlighting how causality is commonly seen as involving more than the merely physical. Consider that it is common to speak of a person being dehydrated resulting in their feeling thirsty ultimately leading to them seeking out something to quench their thirst. The dehydration, a physico-biochemical phenomenon in one domain, causes the subjective phenomenal experience of thirst, a phenomenon belonging to the domain of mental phenomena,[146] that subsequently causes the physico-biochemical effect (in the original, or perhaps a third domain) of seeking out and consuming water, for example. Indeed scientific explanations similarly invoke causal chains that leave and reenter different domains without incurring the wrath of philosophers who charge them with the

[143] Bunge 1977, 503.
[144] Ibid.
[145] I say "all but necessary" because, as already discussed (see Chapter 2), emergent phenomena not only arise between domains but within the same ontological domain as well.
[146] Any distinction between different domains is likely arbitrary. Many emergentists have noted that it is doubtful we will discover any sort of sharp boundaries between different domains (or "levels") of phenomena. See, Humphreys 1997 and Bunge 1980, for more details.

violation of the causal closure of the physical domain. And this is because scientists are not violating the causal closure of the physical domain, rather they are recognizing, as is reflected in their language, that causal powers are not restricted only to fundamentally physical phenomena in our universe. For example, evolutionary biologists commonly connect environmental pressures (physical causes) to organismal fitness (itself a nonphysical effect and cause) to the presence or absence of physiological traits (physical effect).[147]

So if it is the case, as it already appears to be with ordinary language and scientific explanations, that it is common to allow causal chains to leave and reenter different ontological domains, why does downward causation in general, and mental phenomena in particular, apparently present such a difficult problem for the emergentist? Part of the reason is that emergence is often critiqued using metaphysical assumptions no emergentist should hold. One of those assumptions is a commitment to the causal closure of the physical domain. Another is a an acceptance of ontological minimalism. In contrast the emergentist ought to maintain the causal completeness of the physical domain and reject ontological minimalism. Together these metaphysical considerations imply the irreducibility that is one aspect of emergence and a necessary characteristic of ontologically emergent phenomena. Attributing irreducible causal efficacy to mental phenomena, as emergentists do, is therefore only a serious problem for those who wish to simultaneously remain committed to ontological reductionism and minimalism. Indeed a commitment to ontological reductionism and minimalism has serious ontological consequences that the emergentist can avoid. Humphreys for example recognizes that the

[147] See Figure A1 in the Appendix, adapted from Dalziel et al., 2009, which beautifully illustrates the way in which causal chains flow from one domain to another and back again. It would be oversimplifying the causal story, and frankly naive, to insist on reducing all of the interactions highlighted in Figure A1 to causal interaction between fundamentally physical constituents.

generality of both the causal exclusion[148] and downward causation[149] arguments has the potential to empty our world of causation altogether.

> If the exclusion argument does generalize to such hierarchies, and if, for example, chemical and biological events occupy higher levels than do physical events, then no chemical or biological event could ever causally influence a physical event, and if both arguments [the causal exclusion and downward causation arguments] so generalize, then nonreductive physicalism leads to inconsistencies when applied to the general realm of the natural sciences too. The situation is in fact more extreme than this, because most of our physical ontology lies above the most fundamental level, and in consequence only the most basic physical properties can be causally efficacious if these arguments are correct. Indeed, unless we have already isolated at least some of the most fundamental physical properties, every single one of our causal claims within contemporary physics is false and consequently there are at present no true physical explanations that are grounded in causes.[150]

If emergence, as a type of nonreductive physicalism, leads to inconsistencies vis-a-vis the causal exclusion and downward causation problems, it is only because the claims in both arguments are overly simplistic. The advantage of terms like ‘emergent’ and the concept of “emergence” is that they convey the idea that a certain phenomenon is “more than” its fundamental constituents, that at a certain fundamental level, some phenomena are simply ontologically different in kind from other directly related phenomena. The causal exclusion and downward causation arguments simply do not capture the nuances inherent in the causal evolution of natural systems, especially those systems that manifest emergent phenomena. For example, the physical determinism claim of the exclusion argument, that “for every physical event *y*, some

[148] From the claims of (1) exclusion, (2) physical determinism and (3) mind-body dualism, it follows that (4) all mental phenomena are epiphenomena.
[149] From the claims that emergent phenomena (E) only appear when their emergent base (P) is present, and that in a causal chain the next emergent (E*) is caused to appear by the previous emergent (E) causing the next emergent’s (E*) emergent base (P*) to appear, it follows that emergence entails downward causation.
[150] Humphreys 1997a, 3-4.

physical event *x* is causally sufficient for *y*," is far too general to capture notions of causality that apply to sufficiently large and complex systems.[151] This claim (physical determinism) is readily and accurately applicable to describe the movement of billiard balls following the break, which is a relatively simple Newtonian mechanical system. The physical determinism claim utterly fails however in accounting for the cracking and deforming of a concrete sidewalk that occurs as a tree in close proximity to it grows. Presumably the causal chain that weaves through the complex system just described would involve physical, chemical and biological events that together result in the observed physical phenomenon of the concrete sidewalk cracking and moving.

Most of the claims in the causal exclusion and downward causation arguments are, like the physical determinism claim, overly simplistic. The exclusion claim, that "if an event *x* is causally sufficient for an event *y*, then no event *x** distinct from *x* is causally relevant to *y*," equates the ideas of sufficiency, relevance and causation.[152] Far from being interchangeable, these terms imply different states of affairs in regards to the causal relatedness of two phenomena.

> Remember Archimedes' excited outburst on discovering the principle of displacement in his bath. Assuming that his shouting "Eureka!!" was causally sufficient for his cat's startled flight, nobody would think that this disqualified his (simply) shouting from being causally relevant as well. And it would be incredible to treat Socrates' *drinking* the poison as irrelevant to his death, on the ground that his *guzzling* it was causally sufficient...Notice some important differences between causal relevance and sufficiency, on the one hand, and causation, on the other: *x* can be causally sufficient for *y* even though it incorporates enormous amounts of causally extraneous detail, and it can be causally relevant to *y* even though it omits factors critical to *y*'s occurrence. What distinguishes causation from these other relations is that causes are expected to be

[151] Yablo 1992, 247.
[152] Ibid.

> *commensurate* with their effects: roughly, they should incorporate a good deal of causally important material but not too much that is causally unimportant.[153]

As Yablo points out, the concepts of causal sufficiency and causal relevance are significantly different from causation per se. Consider a causal narrative concerning the completion of this thesis. My thoughtful and rapid typing out of this thesis project is causally sufficient for its completion, however both my thoughtfulness and rapidity are causally unnecessary for my typing it to completion.[154] Similarly my tossing of sunflower seeds into the garden is causally relevant for the growth of sunflowers in my garden, but this information alone is causally vague. The cause of there being sunflowers in my garden is not simply that I threw sunflower seeds in there, but also that it was the correct time of year, soil conditions were favourable for their growth, there were adequate biotic and abiotic factors to facilitate their growth, etc. These considerations suffice to show that claims in the causal exclusion and downward causation arguments are far too simple and presupposes that causal antecedents compete with one another as if in a zero-sum game. The result is that only conceiving of the relatedness of two phenomena in a reductively supervening relation obscures important differences between causal sufficiency, relevance and causation per se. Sometimes the relatedness of two phenomena is not a kind of reductive relation however, and in these cases, the concept of emergence is better suited for capturing and signifying the complex causal network responsible for the appearance of an ontologically emergent phenomenon. It is simply not the case that causes compete as if in a zero-sum game: "rather than competing for causal honours, determinables [causes] and their determinates [effects] seem likelier to share in one another's

[153] Ibid., 272-274.

[154] Certainly it would be a poor thesis project and take far longer than my supervisor would like to complete if I was thoughtless and tardy, but these factors have no bearing on whether the thesis is completed per se.

success."[155] This is in fact exactly what happens in cases of emergent phenomena, that is, the novel causal powers of emergent phenomena arise holistically from the causal powers in the emergent base.[156] The implications for the emergentist on the causal exclusion and downward causation problems are clear. In a rather anticlimactic fashion, both problems just dissolve.

> Unless an arbitrary exception is to be made of them, it [the causal exclusion argument] is no argument at all for the causal irrelevance of, say, a sensation that its occurring in some specific physical way was causally sufficient. With events as with properties, physical determinates cannot defeat the causal pretensions of their mental determinables.[157]

Contra Kim and proponents of reduction via supervenience who endorse the causal exclusion and downward causation arguments, it is reasonable to reject the causal closure of the physical domain and thereby reject the conclusions that all mental phenomena are epiphenomenon and that emergence always entails downward causation. Although it may unduly trouble the reductive physicalist, for the emergentist it is simply par for the course to maintain only the causal completeness of the physical domain. In the next chapter I will examine more closely the connection between the novel causal powers of emergent phenomena and how we can represent their existence using the resources of dynamical systems theory.

---

[155] Yablo 1992, 272.

[156] This point will be further explored in the next chapter linking dynamical systems theory (DST) to emergent phenomena. In short, we can represent emergent phenomena using DST. In these dynamical models the causal powers of the emergent phenomenon arise as a result of a fusion of two smaller dynamical systems whose fused dynamical laws determine the evolution of the post-fusion system's state space, i.e., visual representation of the system's possible behaviours. The emergence of life from abiotic chemicals and emergence of complex properties that living organisms possess (e.g., being able to walk), can be accurately and easily modelled using DST by specifically fusing a dynamical model of the environment and a dynamical model of the organism itself together. See, Beer 1995, for more details.

[157] Yablo 1992, 272-273.

Chapter 4

Merging the Old with the New: Emergence and Dynamical Systems Theory

Thus far we have established that ontologically emergent phenomena are characterized by their relationality, novelty, irreducibility and broken symmetry. Moreover the concept of emergence is free from the problems of causal exclusion and downward causation that beset ontologically reductive accounts of phenomena, such as, for example, supervenient accounts of the relationship between mind and body. Although emergent phenomena are neither unexplainable nor unpredictable, the early British Emergentists were correct in thinking that ontologically emergent phenomena are in fact more widespread than might be commonly thought. Protein functionality for example,[158] is an ontologically emergent phenomenon with respect to the biochemical phenomena of living organisms. Life itself is an ontologically emergent phenomenon with respect to abiotic physicochemical phenomena.[159] These examples serve to highlight an advantage of this updated concept of emergence in comparison to the concept as conceived by the early British Emergentists, namely that the latter relied on a hierarchical conception of phenomena within the world whereas the former does not. This is advantageous because any division of phenomena in the world into distinct “levels” is ultimately arbitrary. To maintain that emergent phenomena are always in some sense on a “higher level” than their emergent base is to perpetuate a reductive/constructive view of phenomena that can oversimplify the immense complexity of natural phenomena. Beyond issues stemming from identifying different

[158] See the section titled *Emergence: Broken Symmetries* in Chapter 2.
[159] Ibid.

ontological "levels" it is demonstrably false that emergence entails a relation between "higher level" and "lower level" phenomena. The ontological emergence of superconductivity is an example of an emergent phenomenon that does not arise from a "lower level" phenomenon, rather it exists on the same fundamentally physical level.[160]

The concept of emergence can therefore be understood within a rich conceptual framework. But this is not enough. The concept of emergence advanced here is also put forward as an empirical hypothesis. That is, the concept of emergence is not merely useful as a philosophical heuristic for understanding phenomena in the world. Rather, the concept of emergence can also be cashed out in mathematically precise terms in such a way that emergent phenomena can be explainable, modellable and even in some senses predictable. In this last chapter I argue that the resources in dynamical systems theory (DST) are able to accurately model emergent phenomena so that they become empirically evaluable. In short, if we can model emergent phenomena using DST, then we can compare observations of the phenomenon itself with the modeled phenomenon, and thereby evaluate the phenomenon's status as emergent or not. (This assumes, of course, that we have an accurate and precise model, otherwise disparities between the observed and modeled phenomenon may indicate that a better or more refined model is needed). Indeed the connection between DST and emergent phenomena is commonplace. Tim van Gelder in his paper *The Dynamical Hypothesis in Cognitive Science* notes that limb coordination, for example, is best explained by DST and ought to be considered an "emergent property of a nonlinear dynamical system."[161] Moreover, van Gelder

[160] See the section titled *A Case Study: The Emergence of Superconductivity* in Chapter 2.
[161] van Gelder 1998, 616.

argues that cognition itself, i.e., mental phenomena,[162] are best understood according to the dynamical hypothesis, which is to say that we ought to understand agents that possess mental phenomena (cognitive agents) as dynamical systems. This chapter will build on van Gelder's implicit claim that mental phenomena are best understood according to DST by systemically connecting necessary characteristics of ontologically emergent phenomena to dynamical systems. Further, whereas van Gelder was focused on mind, I will be connecting emergent phenomena in general to DST. Lastly, this chapter will address the historic presumption, perpetuated to this day, that emergent phenomena are unexplainable and unpredictable. I will argue that, as long as we recognize the distinction between ontological and epistemological emergence, the ontological irreducibility of an emergent phenomenon is compatible with its epistemic reducibility. I will attempt to motivate this conclusion by first arguing that emergent phenomena are explainable within the framework of DST, and further that dynamical explanations, i.e., explanations within the framework of DST, conform to the deductive-nomological (DN) model of explanation proposed by Hempel and Oppenheim (H&O).[163] It follows that emergent phenomena, as phenomena explainable by H&O's DN model of explanation, are reductively explainable and predictable, albeit predictable in a certain sense.

[162] Admittedly, van Gelder laments that the concept "cognitive" "resists capture in terms of any concise set of strict conditions." While traditionally to be "cognitive" or considered a "cognitive agent" was associated with the concept of knowledge or "knowledge-based processes," the concept of cognition has grown to include concepts like "intelligence," "adaptability" and "coordination." See, van Gelder 1998, for more details.

[163] See Chapter 1 for more details on H&O's DN model of explanation.

Dynamical Systems

According to DST we can think of a system, epistemically,[164] in the broadest sense as a set of interdependent variables. This interdependence means that changes in any one variable depends on the other variables, and change in those other variables similarly depends on the original variable of interest.[165] The state of a system is therefore the state or numerical value of all the variables at a particular time, all of which exists within a larger state space, which represents the range all of the possible states a system can attain. We can understand the behaviour of a system as the transitions between different states. Unfortunately, as van Gelder notes, "there is no established consensus over what dynamical systems are" and as a result there are a wide range of candidates.[166] From narrow definitions concerning the motion of particles governed by forces to broad definitions that simply subsume all accounts of a "system," it will suffice for us to use the guiding idea of "change in time"[167] as paradigmatic of dynamical systems given that our goal is simply to connect DST to the concept of emergence at a general level.[168] Dynamical systems are therefore systems whose variables change numerically, and hence quantitatively, in time. This property of dynamical systems, of being quantitative, is vitally important because it permits the empirical evaluation of phenomena modelled by DST. Moreover, the quantitative nature of dynamical systems supports a geometric perspective on system behaviour, i.e., we can understand the temporal evolution of a system through its state

[164] That is, the system itself is not merely the interdependent variables, rather these interdependent variables are epistemic entities used to model the system itself.
[165] van Gelder 1998, 616.
[166] Ibid., 617-618.
[167] Ibid., 618.
[168] For some examples of common definitions of dynamical systems, see Table A1 in the Appendix.

space visually, which is "one of the hallmarks of a dynamical orientation."[169] Important terms associated with this geometric perspective that will be necessary to use in our discussion of the connection between DST and emergence include 'state space', 'trajectory,' 'attractor,' 'repellor,' and 'bifurcation.' First, to illustrate the use of these terms in a simple dynamical system, let us consider the simple system of a swinging pendulum, such as one that might be found in a pendulum clock.

In DST, the state space is not merely metaphorical, rather it represents all of the possible behaviours a system may exhibit. The trajectory that a system traces through its state space corresponds to the system's behaviour as it transitions through various states. If our system is an idealized[170] pendulum swinging then it will trace a concentric elliptical trajectory through its state space corresponding to its perpetual oscillating behaviour, as illustrated in Figure 1. The three trajectories represent three pendulums with varying rod lengths as their corresponding positions (y-axis) and momentum (x-axis) change. If however our system is a non-idealized pendulum swinging, then its trajectory through its state space will look quite different. Figure 2 highlights the difference between pendulum behaviour in an idealized system (dotted line) and pendulum behaviour in a non-idealized system (solid line). The behaviour of the non-idealized pendulum varies significantly from the behaviour of the idealized pendulum, and this is represented by the difference in the trajectories traced through the state space of the system. Whereas the idealized pendulum maintains its behaviour indefinitely, the non-idealized pendulum will eventually come to rest as the system loses energy to frictional forces. This is

[169] van Gelder 1998, 619.
[170] By "idealized" I mean that this is a closed system, i.e., there are no frictional or other forces that add or subtract energy from our system.

represented visually by the inward spiraling trajectory of the non-idealized pendulum. In fact any non-idealized pendulum will exhibit the same behaviour, i.e., it will eventually come to rest.

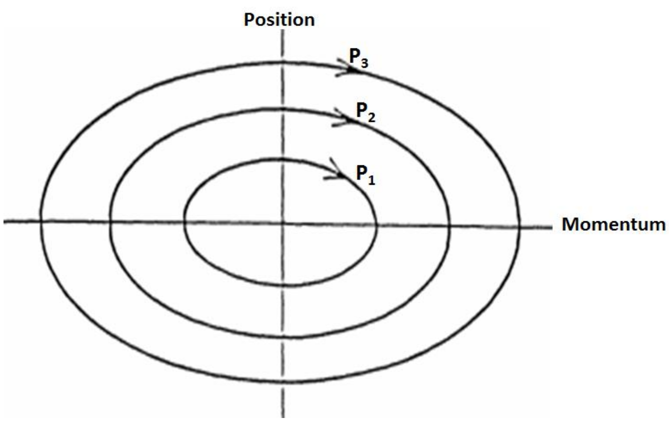

Figure 1: Visualization of the dynamical system representing an idealized pendulum swinging indefinitely. The three concentric ellipses represent three different pendulums (P_1, P_2 and P_3) that possess varying rod lengths, the longest corresponding to the outermost trajectory and the shortest corresponding to the innermost trajectory. Adapted from Rueger 2000, 302.

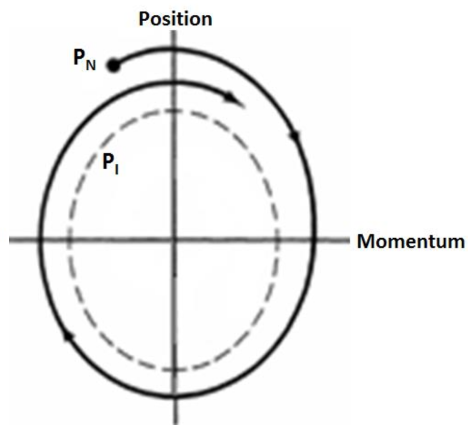

Figure 2: Visualization of two dynamical systems corresponding to an idealized pendulum (P_I)and a non-idealized pendulum (P_N). The trajectories, represent the radically different behaviour between the idealized (dotted line) and non-idealized (solid line) systems. Adapted from Rueger 2000, 302.

In DST terms, we can describe this point in the state space that corresponds to zero momentum and the pendulum's resting position as an "attractor." In short, all dynamical systems that model non-idealized pendulum behaviour will have a point in their state space, an attractor, which the trajectory will eventually reach. In general, an attractor is a point in the state space of a dynamical system that the system will tend toward so long as the system is within what is called the "basin of attraction," i.e., under the influence of an attractor. "Repellors," in contrast, and as the term suggests, are points in the state space of a dynamical system that the system will move away from. Lastly, in DST a "bifurcation" refers to a critical point in the evolving behaviour of a system such that the behaviour pre-bifurcation is radically qualitatively and/or quantitatively different from the behaviour post-bifurcation. Accompanying changes to the state space of a system as a result of a bifurcation include, but are not limited to, the disappearance and appearance of attractors and repellors in the state space. A bifurcation would occur, for example, if we introduced friction into the idealized pendulum system and thereby turn it into a non-idealized system. This small change drastically alters the trajectory of the system through its state space because changing the friction variable introduces an attractor that was not present in the state space of the idealized pendulum system. First order phase transitions, such as the transition of liquid water to solid ice, are also represented as a bifurcation in DST. A small change in the temperature variable, from positive one degree Celsius to negative one degree Celsius for example, corresponds to radical qualitative and quantitative changes in the system.[171]

One important caveat that bears mentioning before proceeding with a discussion of the relation between DST and emergent phenomena is that models are not equivalent to the natural phenomena that they model or purport to explain.

> Models of such phenomena as ferromagnetism, superconductivity, and superfluidity are, like all models, idealized in certain ways. Although there are detailed consequences of these models that fit well with experimental data, the internal structure of the models themselves do not fully represent the structures of the real systems, and in many cases they are known to be gross oversimplifications of the real systems' structure. This is a common feature of models, and because of it we must be very careful indeed when extending claims about the existence of emergence in models to claims about its existence in the systems being modeled.[172]

This gap between models and the phenomena they model is difficult to bridge and makes making claims about ontologically emergent phenomena, insofar as we are able to model ontologically emergent phenomena, inescapably tentative. Nevertheless, in the next two sections I will argue that there are compelling reasons to believe that even if emergent

[171] Quantitatively, solid ice has a larger volume and is less dense than liquid water. Qualitatively, ice is solid and rigid in comparison to the fluidity of liquid water.
[172] Humphreys 2016, 9.

phenomena can be modeled using DST and reductively explained (as well as predicted) according to the DN scheme of explanation, the coherence of the concept of emergence can be defended.

DST and Explanation

Emergent phenomena were historically seen as unexplainable brute facts, but this is not true. If emergent phenomena are modellable using DST, then emergent phenomena may be explainable according to the DN model of scientific explanation proposed by H&O.[173] Contra the early British Emergentists, and contra H&O, emergent phenomena are neither unexplainable nor is ‘emergence’ a term used to simply denote our ignorance surrounding a particular phenomenon. Emergent phenomena are qualitatively different from the phenomena from which they emerge (the novelty criterion), and this qualitative difference is captured in DST and dynamical models of (potentially) emergent phenomena. One way DST captures the existence of ontologically emergent phenomena is via bifurcations which, within a system, signal a radical qualitative shift in the state of the system. Recall the example systems of the idealized and non-idealized pendulums raised earlier. Although not an emergent phenomenon, introducing friction into the idealized pendulum system creates a bifurcation. That is, introducing friction into the idealized pendulum system causes a radical qualitative shift in the system, to wit, the pendulum ceases to move indefinitely and, as a result of the appearance of an attractor in the state space of the system (following the introduction of friction), eventually comes to rest. Importantly, whether modeling non-emergent or emergent phenomena, bifurcations occur when viewing the

[173] Hempel and Oppenheim 1948.

evolution of a system over time. In short, the emergent novelty of a phenomenon, although it may be understood synchronically, is best understood diachronically. Alexander Rueger explains this connection between DST, temporal evolution and potentially emergent phenomena.

> We can say that the behaviour of the system at a time is emergent with respect to the system at an earlier time if some parameter in the base has changed its value slightly during the time interval and the later behaviour is 'novel' compared to the behaviour of the old system and irreducible to it. This is evolutionary or diachronic emergence of properties. The general strategy in developing this notion within the framework of dynamical systems theory is to compare two systems which are connected through a small change in the base properties such that the 'old' system is an unperturbed version of the 'new,' perturbed system, and find out whether this *quantitatively* small perturbation of the base leads to *qualitatively* new behavioural properties.[174]

Recall the example of the ontologically emergent phenomenon of superconductivity.[175] If we compare two potentially superconducting systems, one just above the critical temperature (i.e., the old unperturbed system) and one just below the critical temperature (i.e., the new perturbed system), we find that a small quantitative change in temperature (often a change of one degree Kelvin is sufficient) leads to a qualitatively new property, viz., the appearance of the property of superconductivity or, equivalently, the property of having zero electrical resistivity. On this account, superconductivity, an ontologically emergent phenomenon, is not unexplainable. Superconductivity is not simply a brute empirical fact in the same way that the speed of light in a vacuum is, for example. There are clear material and configurational, in addition to environmental, requirements that must be met before superconductivity can arise. As such superconductivity is neither a brute empirical fact that admits of no explanation (contra early British Emergentists) nor is it out of ignorance that we call superconductivity "emergent"

[174] Rueger 2000, 300.
[175] See the section titled *A Case Study: The Emergence of Superconductivity* in Chapter 2.

(contra H&O, who maintain that "emergence of a characteristic is not an ontological trait inherent in some phenomenon; rather it is indicative of the scope of our knowledge at a given time").[176] Moreover, if it is possible to model the phenomenon of superconductivity using DST, then such a model would align with the DN model of explanation proposed by H&O. That being the case, emergent phenomena would be very much scientifically explainable, a la H&O's DN model of scientific explanation (also known today as the "covering law" model of explanation). Let us examine the connection between dynamical explanations given within the framework of DST and DN style explanations.

According to H&O, to give a DN style explanation (and prediction, since H&O saw them as logically symmetrical) of a phenomenon is to provide the statements that together constitute the explanans. Specifically, these statements include the initial conditions of a system and the general laws that govern the system which together are adduced to account for the phenomenon under investigation.[177] If explanations in DST make use of certain specified initial conditions together with certain specific laws to account for the evolution and change in behaviour of a dynamic system, then dynamical systems explanations conform to the same style of DN explanation proposed by H&O. If DST can model, i.e., explain, emergent phenomena, then it follows that emergent phenomena are scientifically explainable insofar as H&O understand explanation.[178] Prima facie, there is reason to believe that dynamical systems explanations conform to the DN model of explanation. Consider that dynamical systems are characterized by a set of state variables and a dynamical law governing how the values of those state variables

[176] Hempel and Oppenheim 1948, 21.
[177] Ibid., 10.
[178] For the moment I will ignore prediction. Although H&O see explanation and prediction as logically symmetrical, there are reasons to believe that this ought not be the case (see Chapter 1 for more details).

change over time.[179] The state variables and dynamical law are analogous to the initial conditions and general laws that H&O specify ought to appear in the explanans of a DN scientific explanation. In short, to explain a phenomenon using DST is to explain it deductive-nomologically in the way proposed by H&O, a conclusion I hope to motivate given the following considerations.

Joel Walmsley makes two interesting observations about the connection between dynamical models and the DN style of explanation: i) existing dynamical models conform to the DN explanatory scheme advanced by H&O, and ii) proponents of DST explicitly express that it is a goal of DST to discover the law-like ways in which a given system evolves over time. Let us examine these claims in the context of the HKB (Haken-Kelso-Bunz) representation (i.e., dynamical model) of finger movement.

> The HKB model is based upon the simple observation that, when asked to place their hands palm-down and oscillate both index fingers back and forth with the same frequency, people are reliably and stably able to reproduce only two basic patterns. One is where the left index finger and right index finger both move to the left or to the right at the same time [in-phase motion]. The other is where one finger moves to the left, while the other moves to the right, or vice versa [antiphase motion]. Quantifying this observation we can say that the finger movements of subjects are stable when the relative phase of the finger is either 180 degrees (for in-phase motion) or 0 degrees (for antiphase).[180]

What is interesting about this observation of finger movement is that there is an upper bound on the frequency with which a person can sustain the antiphase motion. In short, if a person is told to move their fingers in an antiphase pattern while keeping time with a metronome that is slowly increasing in frequency, a critical point will be reached beyond which only an in-phase

[179] Beer 1995, 176.
[180] Walmsley 2008, 333.

finger motion will be possible to sustain. The fact that above a certain frequency only an in-phase pattern of finger movement is sustainable and stable is an epistemically emergent phenomenon. Indeed Haken et al. note that any speculation on the origin of this in-phase coupling is premature at best and that further research is needed to understand why there are only two stable patterns of finger movement in the first place.[181]

Yet despite a lack of understanding of the origin of the coupling in finger movement or the origin of the stability of only the in-phase movement above a critical frequency, we nevertheless can explain the change in finger movement according to the DN model of explanation. If the explanandum, the phenomenon to be explained, is the switching from antiphase finger movement to in-phase movement following an increase in movement frequency, then the explanans would include the following: statements about the initial conditions (which are necessary in a DN explanation) including the original antiphase motion as well as the starting frequency of finger motion, and statements about the general laws governing the evolution of finger motion (also necessary in a DN explanation), specifically the dynamical law[182] that specifies how changes in frequency are related to changes in coordinated finger movement phase. From statements of the initial conditions and general laws we may deductively derive, a la H&O's DN model of explanation, the phenomenon of interest, i.e., the fact that above a critical frequency only in-phase finger movements are possible to sustain. In the language of DST, there exists two attractors in the state space of the system when the frequency is sufficiently low. Either finger movement will tend towards the stable antiphase motion or it will tend towards the stable in-phase motion. However as the frequency of motion

[181] Haken et al. 1985, 355.
[182] For the dynamical law that represents the coordinated finger motion see Equation A1 in the Appendix.

increases, the system reaches a bifurcation at a critical frequency beyond which only one attractor exists in the state space of the system, namely the in-phase attractor.

Beyond the fact that explanations invoked in DST appear to be constructed in a way that mirrors the DN model of explanation, it is in fact a professed goal of proponents of DST that covering laws, corresponding to the nomological component of the DN model, are sought in order to understand the evolution of a system. Further, dynamical laws are invoked in dynamical explanations to both explain a phenomenon as well as predict future states of a system based on its current state, as van Gelder explains.

> In studying and explaining the behaviour of dynamical systems one aims at formulating equations which describe the evolution of the system, and which can consequently be used to explain why[183] the system is in the state it is in, or to predict what states it will come to be in...If we know the current state of the system, i.e., the point in state space it currently occupies - then we can use the equations governing the behaviour of the system to determine what point it will occupy next.[184]

Given that H&O maintain that an "event under discussion is explained by subsuming it under general laws," coupled with the fact that it is the explicitly expressed goal of dynamicists to formulate equations, i.e., general laws, which describe the evolution of a system, it follows that dynamical explanations explain phenomena in just the way articulated by H&O.[185] Dynamical explanations, i.e., explanations in DST, just are DN style explanations. This is further supported by the fact that, in accordance with the DN model of explanation, dynamical explanations of phenomena "unfold in exactly the way described by the rule [i.e., dynamical law]" governing the

---

[183] Recall that answering the "why?" question in regards to a particular phenomenon is the quintessentially H&O idea of scientific explanation.
[184] van Gelder 1991, 500.
[185] Hempel and Oppenheim 1948, 10.

behaviour of the phenomena which allows for its deductive derivability.[186] This deductive aspect of the DN model of explanation is necessary to understand how a phenomenon is both explained as well as predictable from initial conditions and general laws. Bearing in mind the idea that dynamical systems are characterized by change in time, van Gelder and Port elaborate on the deductive/deterministic nature of dynamical explanations writing:

> Now, dynamical models based on differential equations are the preeminent mathematical framework science uses to describe how things happen in time. Such models specify how change in state variables at any instant depends on the current values of those variables themselves and on other parameters. Solutions to the governing equations tell you the state that the system will be in at any point in time, as long as the starting state and the amount of elapsed time are known.[187]

To reiterate, since the state of a dynamical system can be mathematically calculated given the initial conditions, i.e., the "starting state" and "amount of time elapsed," and laws governing the system's behaviour, i.e., the "governing equations," then dynamical explanations given within the epistemic framework of DST explain phenomena exactly according to the DN scheme H&O propose.

A final consideration that should motivate an acceptance of explanations in DST as DN style explanations is that dynamical explanations satisfy H&O's necessary logical and empirical conditions of adequacy for DN explanations. Dynamical explanations satisfy the first logical condition that the explanandum must be a "logical consequence" or "logically deducible" from the explanans.[188] Indeed the kind of explanation invoked in DST to describe dynamical systems relies on the logical deducibility of the state of a system, the explanandum, from the information

[186] van Gelder and Port 1995, 14.
[187] Ibid., 19.
[188] Hempel and Oppenheim 1948, 11.

contained in the initial conditions and dynamical laws, the explanans. Given that we have already noted that DST makes use of dynamical laws and that these laws are actually required for the derivation of the state of a system at a particular time, H&O's second logical condition necessary for DN explanations has also been met.[189] The third logical condition necessary for DN explanations that H&O insist on is that the explanans must have empirical content, and this is obviously the case for explanations in DST if we consider the HKB model of finger movement given that, as Walmsley points out, "it was, after all, discovered on the basis of observations of rhythmic finger movement in human subjects."[190] So dynamical explanations offered in the epistemic framework of DST fall under the general scheme of DN models of explanation proposed by H&O. Consequently emergent phenomena modellable via DST are scientifically explainable.

DST and Prediction

Thus far I have avoided a discussion of the relation between explanations and predictions in DST. If we are correct in claiming that dynamical systems explanations conform to H&O's DN model of explanation, then it must also be the case, via the symmetry thesis, that phenomena explained by DST are also in principle predictable. Recall H&O's insistence that considerations surrounding their account of explanation apply equally to prediction. In short, if the explanans is sufficient to deduce the explanandum after the observation of a particular phenomenon (i.e., explanation) then it is equally sufficient to deduce that a particular phenomenon will occur prior to its

[189] Ibid.
[190] Walmsley 2008, 341.

observation (i.e., prediction). Just as emergent phenomena were historically seen as unexplainable brute facts, they were similarly seen and continue to be associated with the idea of unpredictability. However similar to the purported unexplainability of emergent phenomena, absolute unpredictability has been erroneously associated with the concept of emergence.

Given the symmetry between explanation and prediction on H&O's DN model of explanation, together with the fact that DST can model emergent phenomena, it follows that emergent phenomena are both explainable and predictable. More than unexplainability, unpredictability has been closely linked to the concept of emergence. There is a tension therefore between DN explanations of emergent phenomena and the tendency to conceive of emergent phenomena as inherently unpredictable. This tension is all the more evident when we model cognitive systems using DST or apply it more broadly in the cognitive sciences.

> Since covering law explanations [or synonymously, deductive-nomological explanations] require *deducibility* of the explanandum from the explanans, whilst most conceptions of emergence (as a non-reductive position) require the *absence* of such deducibility, it seems that dynamical cognitive science is in direct *conflict* with emergentism about the mind. Given that covering law explanation is often seen as a good example of *reductive* explanation, behaviours which can be explained dynamically cannot be regarded as "emergent" in any non-reductive sense.[191]

The basic idea is that the coherence of emergence, as one version of nonreductive physicalism, is threatened if emergent phenomena are in principle predictable, a fact which follows from the reductive explanation of emergent phenomena (insofar as those phenomena are explained according to the DN model of explanation). In short, the concern is that the deducibility of explanations in DST that purport to model emergent phenomena precludes our ability to

[191] Ibid., 346.

understand them as emergent phenomena, since emergent phenomena are supposedly unpredictable. Accepting that emergent phenomena can be explained according to the DN model of explanation also means accepting that emergent phenomena are both predictable and reductively explainable, claims that have historically been eschewed by emergentists. Nevertheless I maintain that notions of predictability and reductive explainability are compatible with the account of emergence advanced in this thesis.

If we were interested in requiring the epistemic irreducibility and in-principle unpredictability of emergent phenomena, there may be ways to rescue the coherence of emergence. Walmsley for example suggests that we could save the intuition that DST can model emergent phenomena, specifically the mind, if we reconstrue the concept of emergence so that it is a thesis about laws only, instead of a more robust ontological thesis[192] about laws, objects and properties.[193] Such a reconstruction would allow us to insist on the irreducibility and unpredictability of these emergent or "dynamical laws" from other "lower level" laws and so preserve the intuitive ideas of epistemic irreducibility and unpredictability associated with the concept of emergence. But we need not require the epistemic irreducibility and unpredictability of emergent phenomena to maintain the coherence of the ontological irreducibility of emergent phenomena. Walmsley for example appears to conflate epistemological emergence with ontological emergence when he argues that behaviours which can be reductively explained cannot thereby be labeled emergent behaviours.[194] However it simply does not follow that if an emergent phenomenon, mind for example, is reductively explainable it is therefore also ontologically reducible. The early British Emergentists similarly conflated an understanding or

[192] Refer to Chapter 2, specifically the section titled *Emergence: Relationality* for more details on the ontological relata that may be related via emergence.
[193] Walmsley 2008, 346.
[194] See Footnote 191.

knowledge of a phenomenon with the phenomenon itself; they did not carefully distinguish ontological from epistemic emergence. A careful examination of the concept of emergence reveals that it has distinct and separate applications in the realms of ontology and epistemology. It is perfectly intelligible to claim that a phenomenon is both ontologically emergent and reductively explainable. This is the case for the ontological emergence of superconductivity. It is also perfectly intelligible to claim that a phenomenon is both ontologically reducible yet epistemically emergent, as would be the case for phenomena like flocking birds, traffic jams or schooling fish. A flock is ontologically reducible to its constituent members, but the concept "flock" is an epistemically irreducible emergent phenomenon. Bunge summarizes this point using water as an example.

> A body of water is a system, hence something with a structure, not only a composition. And that structure includes the hydrogen bonds among H_2O molecules. The result is a system with emergent properties such as fluidity, viscosity, transparency and others, which its molecular components lack. Surely one can (hope to) understand all of these emergent properties in terms of those of the water molecules and their interactions. That is, one can (hope to) 'reduce' the macroscopic properties of water to the properties of its microcomponents. But such an explanation — which has yet to be provided — does not accompany an ontological reduction: explained fluidity is still fluidity. Likewise explained vision is still vision, explained imagination is still imagination, and explained consciousness is still consciousness. Therefore ontological reductionism is just as untenable in the matter of mind as it was found to be in the matter of matter.[195]

So despite the fact that emergent phenomena are explainable according to the DN model via dynamical explanations and hence reductively explainable as well as predictable, this does not undermine the coherence of the concept of ontological emergence.

[195] Bunge 1977, 506.

The notion of unpredictability typically associated with emergent phenomena is a simple binary, all or nothing classification. In short, it has been argued that emergent phenomena are inherently unpredictable phenomena in contrast to non-emergent phenomena which are, at least in principle, predictable. But this is far too simple a notion of unpredictability. Although H&O maintain that any DN style explanation is equally a prediction, this is not always the case in DST, especially in dynamical systems that explain emergent phenomena. A symmetry between explanation and prediction is quite obvious in the case of relatively simple dynamical systems in which the state space and the attractors and repellors within it remain relatively constant. Just as we can explain how a pendulum comes to rest by pointing to the attractor within its state space, so too could we predict that a pendulum recently set in motion will also come to rest given the presence of a similar attractor in its state space. Predictions of hitherto unobserved phenomena are also possible given certain dynamical laws governing the evolution of a system and given stipulated initial conditions. For example, following their observation of rhythmic finger movement and the development of the HKB model, Haken et al. noted that "although the phase transition [from antiphase to in-phase motion] occurred at very different frequencies of hand motion for different subjects, it was nevertheless predictable."[196] Further, the HKB model was able to generate predictions that were subsequently empirically verified. These included predictions of the results of selective interference on finger motion, such as by applying a small electrical current, which were vindicated via observation.[197] In accordance with H&O's articulation of the symmetry between explanation and prediction, the DN form of dynamical explanations allows for the support of counterfactuals, i.e., we can say "how the dynamical

[196] Haken at al. 1985, 347.
[197] Walmsley 2008, 334.

system in question would have behaved in various non-actual circumstances," and consequently is capable of generating predictions.[198] But not all dynamical systems are able to generate such precise predictions.

DST is able to capture a strong idea of unpredictability, albeit a weaker one than inherent unpredictability, that many proponents of emergence insist is characteristic of ontologically emergent phenomena. Clearly most emergent phenomena are predictable in a certain sense ex post facto. Prior to the observation of changes in the phase of finger motion as movement frequency increased, such a change would have been highly unpredictable. Likewise prior to the observation of superconductivity as temperature is lowered, such a fundamental change in a metal's conductive properties was simply not expected or predicted to occur. Following observation however, phase changes in finger motions and changes in electrical resistivity respectively entered the realm of the predictable, at least for sufficiently similar systems.[199] Yet there exist certain systems, modellable by DST, that do no support robust prediction making. Weather systems are an example of ontologically emergent phenomena that are explainable but highly unpredictable.[200] Certainly weather is predictable on small time scales

---

[198] Ibid., 337.

[199] The discovery of high-temperature superconductors, for example, illustrates how predictive capabilities extend only so far. After the discovery of the emergent phenomenon of superconductivity in mercury, many other metals were predicted, successfully, to also possess the property of superconductivity after being cooled below a critical temperature. These systems of "ordinary" or metallic superconductors were similar enough to permit predictions of their superconducting capabilities. This was not the case for high-temperature superconductors whose existence was, again, neither expected nor predicted until they were first observed in the 1980's (superconductivity was first observed in 1911). As the name suggests, high-temperature superconductors are materials whose critical temperatures are significantly higher than ordinary superconductors: they are significantly different systems from ordinary superconductors, and as a result, their existence could not have been predicted before their observation (at least on the basis of ordinary superconducting systems).

[200] Although it will likely require elaboration, a rough argument for the ontological emergence of weather would proceed as follows. Weather is inherently relational, i.e., the formation of a hurricane is determined by air pressure, sea surface temperature, etc. Weather also exhibits

and across small geographical areas, but it becomes increasingly unpredictable on longer time scales and over larger geographical areas. All emergent phenomena share this feature: they are predictable, but this predictive power sharply drops off as a function of system size and temporal removal. In short, predicting an emergent phenomenon is much easier when considering a relatively simple system over a short time span whereas it is effectively impossible to predict the appearance of an emergent phenomenon when considering a relatively complex system as it evolves over long periods of time (e.g., cosmically significant periods of time). This is because the state space of dynamical systems, such as a weather system, is constantly changing resulting in an unpredictable trajectory through the dynamically evolving state space. Additionally, dynamical systems modelling weather was where it was discovered that there exists a wholly different kind of attractor that can appear in dynamical systems: the chaotic attractor.

> In 1963 Edward N. Lorenz of the Massachusetts Institute of Technology discovered a concrete example of a low-dimensional system that displayed complex behaviour. Motivated by the desire to understand the unpredictability of the weather, he began with the equations of motion for fluid flow (the atmosphere can be considered a fluid), and by simplifying them he obtained a system that had just three degrees of freedom [i.e., three variables that specify the state of the system]. Nevertheless, the system behaved in an apparently random fashion that could not be adequately characterized by any of the three attractors then known. The attractor he observed, which is now known as the Lorenz attractor, was the first example of a chaotic, or strange, attractor. Employing a digital computer to simulate his simple model, Lorenz elucidated the basic mechanism responsible for the randomness he observed: microscopic perturbations are amplified to affect macroscopic behaviour. Two [trajectories] with nearby initial conditions diverge exponentially fast and so stay close together for only a short time.

---

novelty in the sense that entities like a tornado are not included under the closure criteria appropriate for the behaviour of gases. Likewise weather is irreducible to gaseous and/or vaporous constituents and breaks continuous translational and rotational symmetry. It has been suggested that weather, tornadoes for example, should be thought of as kinematic manifestations of dynamic processes.

> The situation is qualitatively different for non-chaotic attractors. For these [non-chaotic attractors], nearby [trajectories] stay close to one another, small errors remain bounded and the behaviour is predictable.[201]

Complex and chaotic dynamical systems, along with the discovery of chaotic attractors, illuminated the fact that the evolution of certain systems was unpredictable. More precisely, complex and chaotic dynamical systems illuminated how the predictive powers of certain dynamical models, such as those representing emergent phenomena, have predictive powers that attenuate the further into the future one wishes to make a prediction. Importantly, this is not to say that the evolution of these systems in which emergent phenomena appear is also unexplainable. Ex post facto, the evolution of a weather system for example, could be fully understood on the basis of its observed behaviour. That is, the system can be explained by noting how certain (initial) conditions, together with the dynamical laws specifying the features of the system's state space (i.e., where and what kind of attractors/repellors are present in the state space), entail the observed phenomena.

Based on the above considerations we can rescue the idea that emergent phenomena are unpredictable while simultaneously maintaining that emergent phenomena are explainable. The presence of chaotic attractors within a dynamical system is indicative of the fact that the system is highly unpredictable, especially over long time scales. Nevertheless this does not render unintelligible the fact that the system is understandable and explainable by noting how the properties of the system's state space influenced its trajectory (i.e., the behaviour of the system) through the state space.

---

[201] Crutchfield et al. 2008, 380.

Conclusion

The concept of emergence has an important ontological and epistemological role denoting a commitment to understanding phenomena according to the metaphysical presupposition of nonreductive physicalism. It offers an alternative and superior way to understanding phenomena beyond the reductive/constructive physicalist paradigm that has dominated philosophical thought, especially in the philosophy of mind. Indeed advances in the field of artificial intelligence (AI) have illuminated not only how little we know about certain phenomena, like creativity and intelligence for example, but also how little we know about how these phenomena arise from other directly related phenomena. The creativity and intelligence displayed by AlphaGo, the Go game playing AI, that bested Lee Sedol, one of the most prolific Go players, was both unexpected and unmatched by the human player. The revival of the concept of emergence, especially now in the early 21st century when we stand on the precipice of developing highly intelligent machines, is not coincidental. What the early British Emergentists could only speculate on over a century ago, that mind emerges from neurophysiological processes, can be examined using an ever increasing repertoire of scientific tools. These tools range from computer models and simulations to experiments in the fields of neuroscience, psychology and cognitive science, and the results of decades of experiments have begun to paint the compelling picture of an emergent mind that is but one aspect of an emergent cosmos.

The account of emergence defended in this thesis was that of ontological emergence. In short, I maintain that there exists a plurality of ontological domains, each with their own domain specific laws and phenomena, inextricably linked to other ontological domains. The normative claim maintained throughout this thesis has been that all ontologically emergent phenomena possess four characteristic features: relationality, novelty, irreducibility and broken symmetry. Emergent phenomena always emerge from, and therefore are necessarily directly related to, other phenomena, the emergent base. Perhaps the most striking feature of emergent phenomena is their second characteristic. Emergent phenomena are novel with respect to the phenomena from which they emerge, i.e., they are different in kind from the phenomena that constitute the emergent base. Ontological emergence entails the irreducibility of the emergent phenomenon to its emergent base. Importantly, this is not to say that emergent phenomena are also epistemically irreducible, only that the kind of ontological relation that holds between an emergent phenomenon and its emergent base is fundamentally nonreductive. Emergent phenomena are neither ontologically eliminable, identifiable with, or supervenient upon their emergent base. Finally, emergent phenomena are characterized by a broken symmetry between the emergent phenomenon and the emergent base. Just as emergent phenomena are not reducible to their constituents, neither are they constructible from the constituents of the emergent base. Emergent phenomena exhibit a complexity not exhibited by their emergent base, and this complexity precludes their construction from the mere constituents of the emergent base. Instead this change in complexity manifests as a broken symmetry that accompanies the appearance of an emergent phenomenon.

All of the examples of emergent phenomena offered throughout this thesis possess the four necessary characteristics of emergence, since this is also a descriptive account of candidate ontologically emergent phenomena.[202] Phases of matter for example are emergent with respect to one another. If our emergent base is a system of liquid water, then the appearance of solid ice

---

[202] See Chapter 2 for an in-depth exploration of the emergence of superconductivity from a metallic conductor.

is an emergent phenomenon with respect to its liquid water emergent base. Relationality is satisfied because the ice indeed emerges from the water. Novelty is similarly satisfied considering that solids are different in kind from liquids, e.g., liquids fill the shape of their container whereas solids do not, solids are rigid and liquids are not, etc. Irreducibility is also satisfied given that solid ice is not ontologically reducible to liquid water. Solid ice is not a candidate for elimination from our ontology, does not share a relation of identity with liquid water, and does not supervene on liquid water. Solid ice has an ontic status equal to but ultimately independent of the ontic status of liquid water, each of which belong to different but connected ontological domains (e.g., perhaps the ontological domains of "crystalline solids" and "liquid" respectively). Lastly, since solid ice is emergent with respect to its emergent base of liquid water, such an emergence must exhibit a broken symmetry, and this is indeed the case. The change in complexity from liquid water to solid ice manifests as a breaking of continuous translational and rotational symmetry.

The concept of emergence is attractive in part because it avoids many of the pitfalls associated with other concepts that attempt to make sense of the relation between mind and body in particular and the relatedness of seemingly disparate phenomena in general. It is not the case, as some philosophers have argued, that emergence entails supervenience and thereby entails grappling with the serious problems of causal exclusion and downward causation. If emergence does entail supervenience, then it follows from the conclusions of the causal exclusion argument that all supervening phenomena are epiphenomena (i.e., causally inefficacious), and the downward causation argument that emergent phenomena must possess causal efficacy, and therefore the concept of emergence is dangerously self-contradictory. But this conclusion only arises if one insists on using metaphysical claims that are already in conflict with the concept of emergence. Rather anticlimactically, the problems of causal exclusion and downward causation simply dissolve for the emergentist once the proper metaphysical presuppositions (i.e., a rejection of ontological minimalism and commitment to the causal completeness, not causal closure, of the physical domain) are in place.

Lastly, another strength of the account of emergence advanced here is its compatibility with dynamical systems theory (DST). Although an epistemic tool, DST is an ideal framework through which to understand ontologically emergent phenomena because necessary characteristics of ontological emergence, like novelty and broken symmetry, can be cashed out in the mathematically precise terms of DST. This is of critical importance because the account of emergence being advanced here is in part empirical. That is, the categorization of candidate phenomenon as emergent or not is in part an empirical matter and not merely within the realm of philosophical debate. To be sure, an additional advantage of representing emergent phenomena using DST is that such an epistemic framework dispels the aura of mystery, unexplainability and unpredictability that pervades discussions of ontological emergence. Contrary to what early British Emergentists and critics of the concept of emergence thought, emergent phenomena are not mysterious nor are they unexplainable or unpredictable. In fact if, as I argued,[203] emergent phenomena can be modeled using DST, then dynamical systems explanations conform to the deductive-nomological (DN) scheme of scientific explanation proposed by Hempel and Oppenheim (H&O).[204] This being the case, emergent phenomena are decidedly not unexplainable and neither are they wholly and completely unpredictable. Rather the explanation, and hence predictability (a la H&O's symmetry between explanation and

[203] See Chapter 4.
[204] Hempel and Oppenheim 1948.

prediction) hinges on discovering the dynamical laws appropriate for a given emergent phenomenon. The initial discovery of the emergent phenomenon of superconductivity in mercury for example prompted predictions about the emergence of superconductivity in other conventional metals, predictions which were later vindicated. What was not predicted however was the emergence of superconductivity from wholly different kinds of phenomena, hence the distinction between conventional superconductors and so called high temperature superconductors. So despite the fact that ontologically emergent phenomena are both explainable and predictable, this does not render the concept of emergence incoherent. It merely highlights how our understanding of ontologically emergent phenomena according to the DN model of scientific explanation entails their epistemological reducibility. That is, we can reductively explain an ontologically emergent phenomenon, like mind or superconductivity i.e., the explanandum, by identifying the explanans (i.e., initial conditions and dynamical laws) required to derive the explanandum of interest.

Ultimately my aim has been to provide a positive account of the concept of emergence and to identify potentially ontologically emergent phenomena. Further, I have attempted to separate the concept of emergence from other similar concepts, such as supervenience, and thereby distance emergence from the conceptual problems associated with those concepts, problems such as the causal exclusion and downward causation problem. Moreover, I have argued that despite the fact that our understanding and explanations of ontologically emergent phenomena conform to the DN scheme of explanation advanced by H&O, this does not threaten the integrity of the concept of emergence. Indeed an unfortunate and erroneous presupposition that I hope to have dispelled is the idea that emergent phenomena are wholly unexplainable and unpredictable.

Future Directions

The concept of emergence was introduced partly to make sense of the phenomenon of mind. But this is a broad concept since mind can be taken to subsume any number of other concepts such as intelligence, creativity, information-processing or self-reflective thought. Each facet of the phenomenon of mind may be emergent with respect to certain neurophysiological systems, and recent advances in the field of AI appear to support thinking in such directions. The computer program AlphaGo's victory against world renowned Go player Lee Sedol (4-1 for AlphaGo) highlights how artificial neural networks can approach, and in certain domains surpass, human levels of strategic planning and other mental limits inherent in the biological human mind.[205] I believe that the concept of emergence has enormous potential to track these growing developments in the field of AI. Consider that the emergence of mind from biological carbon-based organisms on the one hand might draw us towards looking for signs of mind in other biological organisms,[206] but may on the other hand render us insensitive in our searches for mind that emerges from artificial silico-metallic entities. Human level AI is decades if not at least a century away, however the advantage of the concept of emergence is its potential ability to measure and quantify growth in the field and the potential power of AI. Self-driving cars for example are artificial entities that possess visual perception, planning and reasoning capabilities as well as reactive planning and reasoning skills (i.e., they do not simply drive based on a previously decided upon plan with no regard for current perceptions), albeit all at a subpar

[205] See Wang et al., 2016 for more details.

[206] And indeed we have been relentless in our search for signs of intelligence, problem solving, a recognition of self, i.e., signs of mental phenomena, in animals ranging from crows and elephants to dogs and cats.

human level.[207] Such AI, I submit, is far stronger than AlphaGo's intelligence, which although powerful, is limited to a very narrow domain: playing the game Go. In other words, the suite of emergent phenomena, primarily mental phenomena, that emerge from the self-driving car system (including its programing but also the car body and driving environment) is indicative of its strength as an AI in comparison to the relative few emergent phenomena, again primarily mental, that emerge from the AlphaGo system.

This is just one potential application of the concept of emergence and a future direction of investigation that may be fruitful. Many other domains and areas of research have the potential to make use of the idea of emergence. Advances in virtual reality (VR), similar to AI, can be understood in an emergentist framework as can human social and political structures. Ultimately, the fate of the concept of emergence will be determined empirically, but at present the concept is, at the very least, a useful stepping stone for understanding the complexity, variety and relatedness of natural phenomena in the ever evolving cosmos.

[207] See Levinson et al., 2011 for more details.

Bibliography

Alexander, Samuel. 1920. *Space, Time, and Deity*. London: Macmillan and Co.

Anderson, Philip W. 1972. “More Is Different.” *American Association for the Advancement of Science,* 177(4047), 393-396.

Anderson, Philip W. 1981. “Some General Thoughts About Broken Symmetry.” *Department of Physics, Princeton University, N.J.*, 1-17.

Anderson, Philip W. 1959. “Theory of Dirty Superconductors.” *Journal of Physics and Chemistry of Solids,* 11, 26-30.

Arthur, Richard. 2014. *Leibniz*. Malden: Polity Press.

Beer, Randall D. 1995. “A dynamical systems perspective on agent-environment interaction.” *Artificial Intelligence,* 72, 173-215.

Brown, Harvey R., and Peter Holland. 2004. “Dynamical vs. variational symmetries: understanding Noether’s first theorem.” *Molecular Physics,* 102(11-12), 1133-1139.

Bunge, Mario. 1977. “Emergence and the Mind.” *Neuroscience,* 2, 501-509.

Bunge, Mario. 1980. *The Mind-Body Problem: A Psychobiological Approach*. Toronto: Pergamon Press.

Crutchfield, James P., J. Doyne Farmer, Norman H. Packard, and Robert S. Shaw. 2008. “Chaos.” In *Emergence: Contemporary Readings in Philosophy and Science*, edited by Mark A. Bedau and Paul Humphreys, 375-386. Cambridge: The MIT Press.

Dalziel, Anne C., Sean M. Rogers, and Patricia, M Schulte. 2009. “Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology.” *Molecular Ecology,* 18, 4997-5017.

Douglas, Heather E. 2009. “Reintroducing Prediction to Explanation.” *Philosophy of Science,* 76(4), 444-463.

Eddington, Arthur. 1968. *The Nature of the Physical World*. Ann Arbor: The University of Michigan Press.

Haken, Hermann, J. A. Scott Kelso, and H. Bunz. 1985. “A Theoretical Model of Phase Transitions in Human Hand Movements.” *Biological Cybernetics,* 51, 347-356.

Hanson, Norwood R. 1959. “On the Symmetry Between Explanation and Prediction.” *The Philosophical Review,* 68(3), 349-358.

Hempel, Carl G., and Paul Oppenheim. 1948. “Studies in the Logic of Explanation.” *Philosophy of Science,* 15(2), 567-579.

Hirsch, Jorge E. 2009. “BCS theory of superconductivity: it is time to question its validity.” *Physica Scripta,* 80(3), 1-13.

Humphreys, Paul. 2016. *Emergence: A Philosophical Account*. New York: Oxford University Press.

Humphreys, Paul. 1997. “Emergence, Not Supervenience.” *Philosophy of Science,* 64, S337-S345.

Humphreys, Paul. 1997a. “How Properties Emerge.” *Philosophy of Science,* 64(1), 1-17.

Kelley, Lawrence A., and Michael J. E. Sternberg. 2009. “Protein structure prediction on the Web: a case study using the Phyre server.” *Nature Protocols,* 4(3), 363-371.

Kim, Jaegwon. 1992. “”Downward Causation” in Emergentism and Nonreductive Physicalism.” In *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, edited by Ansgar Beckermann, Hans Flohr and Jaegwon Kim, 119-138. Berlin: Walter de Gruyter and Co.

Kim, Jaegwon. 1999. “Making Sense of Emergence.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition,* 95(½), 3-36.

Kim, Jaegwon. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge: The MIT Press.

Kim, Jaegwon. 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.

Leggett, Anthony J., and Fernando Sols. 1991. “On the Concept of Spontaneously Broken Gauge Symmetry in Condensed Matter Physics.” *Foundations of Physics,* 21(3), 353-364.

Levinson, Jesse, Jake Askeland, Jan Becker, Jennifer Dolson, David Held, Soeren Kammel, J. Zico Kolter, Dirk Langer, Oliver Pink, Vaughan Pratt, Michael Sokolsky, Ganymed Stanek, David Stavens, Alex Teichman, Moritz Werling, and Sebastian Thrun. 2011. "Towards Fully Autonomous Driving: Systems and Algorithms." *Intelligent Vehicles Symposium (IV) IEEE*, 163-168.

Mastin, L. The Physics of the Universe. "The Universe By Numbers." Accessed August 14, 2018. https://www.physicsoftheuniverse.com/numbers.html

McLaughlin, Brian P. 2008. "The Rise and Fall of British Emergentism." In *Emergence: Contemporary Readings in Philosophy and Science*, edited by Mark A. Bedau and Paul Humphreys, 19-60. Cambridge: The MIT Press.

Morgan, C. Lloyd. 1923. *Emergent Evolution*. London: Williams and Norgate.

Morgan, C. Lloyd. 1912. *Instinct and Experience*. London: Metheun and Co.

Nagel, Ernest. 2008. "Issues in the Logic of Reductive Explanations." In *Emergence: Contemporary Readings in Philosophy and Science*, edited by Mark A. Bedau and Paul Humphreys, 359-374. Cambridge: The MIT Press.

Rueger, Alexander. 2000. "Physical Emergence, Diachronic and Synchronic." *Synthese,* 124(3), 297-322.

Schon, Gerd and A. D. Zaikin. 1990. "Quantum Coherent Effects, Phase Transitions, and the Dissipative Dynamics of Ultra Small Tunnel Junctions." *Physics Reports,* 198(5-6), 237-412.

Scriven, Michael. 1962. "Explanations, Predictions, and Laws." *Minnesota Studies in the Philosophy of Science,* 2, 170-230.

Silberstein, Michael. 2001. "Converging on Emergence." *Journal of Consciousness Studies,* 8(9-10), 61-98.

Teller, Paul. 1992. "A Contemporary Look at Emergence." In *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, edited by Ansgar Beckermann, Hans Flohr and Jaegwon Kim, 139-156. Berlin: Walter de Gruyter and Co.

Van Gelder, Tim. 1991. "Connectionism and Dynamical Explanation." *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society,* 499-503.

Van Gelder, Tim and Robert F. Port. 1995. "It's About Time: An Overview of the Dynamical Approach to Cognition." In *Mind as Motion: Explorations in the Dynamics of Cognition,* edited by Robert F. Port and Tim van Gelder, 1-43. Cambridge: The MIT Press.

Van Gelder, Tim. 1998. "The Dynamical Hypothesis in Cognitive Science." *Behavioral and Brain Sciences,* 21, 615-628.

Van Gulick, Robert. 2001. "Reduction, Emergence and Other Recent Options on the Mind/Body Problem." *Journal of Consciousness Studies,* 8(9-10), 1-34.

Walmsley, Joel. 2008. "Explanation in Dynamical Cognitive Science." *Minds and Machines,* 18, 331-348.

Wang, Fei-Yue, Jun Jason Zhang, Xinhu Zheng, Xiao Wang, Yong Yuan, Xiaoxiao Dai, Jie Zhang, and Liuqing Yang. 2016. "Where Does AlphaGo Go: From Church-Turing Thesis to AlphaGo Thesis and Beyond." *IEEE/CAA Journal of Automatica Sinica,* 3(2), 113-120.

Yablo, Stephen. 1992. "Mental Causation." *The Philosophical Review,* 101(2), 245-280.

Appendix

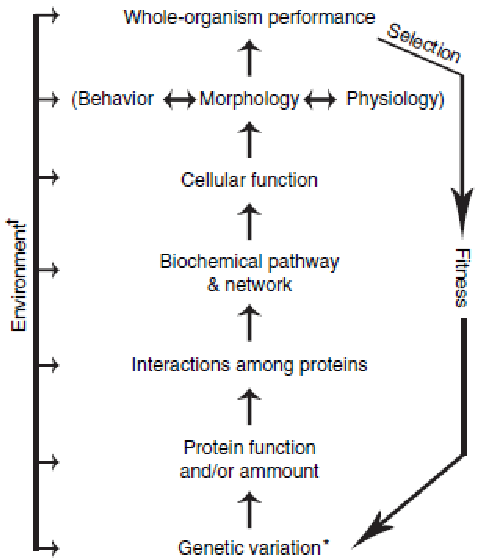

Figure A1: Adapted from Dalziel et al. 2009, this figure highlights how scientists commonly conceive of physical and nonphysical causes interacting.

Guiding idea
1. A system of bodies whose motions are governed by forces. Such systems form the domain of dynamics considered as a branch of classical mechanics.
2. A physical system whose state variables include rates of change.
3. A system of first-order differential equations; equivalently, a vector field on a manifold.
4. Mapping on a metric space.
5. State-determination.
6. Any mapping, equation, or rule.
7. Change in time.

Table A1: There are many different ways to think of dynamical systems, including some listed here. For more details and examples of each kind of system see van Gelder 1998.

$$\frac{d\phi}{dt} = -a \sin \phi - 2b \sin 2\phi$$

Equation A1: This equation represent the dynamical law that governs the relation between the rate of change of relative phase and the periodic function of current relative phase and frequency of oscillation. Both relative phase (Φ) and the frequency of oscillation (inversely proportional to b/a) are captured in this covering law. Adapted from Walmsley 2008.