The Emergence of Life(?)

by

Zachary Munroe

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Philosophy

Waterloo, Ontario, Canada, 2009

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

While emergentism is a frequently debated and contentious topic in some areas of philosophy, it is not discussed as often in the sciences. Where it does appear in scientific literature, it is usually a weak formulation that admits as emergent many entities and properties that would not be considered emergent under a stronger formulation. Premature admission of this sort sometimes occurs in the context of physics, but it is more likely to occur in higher-level sciences like biology. In this thesis, I examine the claim that life, a fundamental biological feature, is emergent. In order to do this, I begin by examining what counts as life. I settle on three features that are necessary for life as we know it, and I show that the smallest unit of life is the prokaryotic cell. I then examine the received view of emergentism and identify its key tenets. I consider two of these in depth, as they play a crucial role in my argument. Finally, I consider weaker formulations of emergence found in scientific literature in order to contrast them with the robust philosophical notion developed within this thesis. I argue that, based on a strong formulation of emergence, life should not be considered emergent, though some may dispute this position if a weaker version of emergence is adopted.

Acknowledgments

This thesis was only made possible by contributions and support from several people. With no attention to order of appearance I need to thank:

Steve Weinstein, my thesis supervisor, whose patience and guidance helped to shape my initial ramblings into something coherent and purposeful.

Doreen Fraser and Paul Thagard, whose comments and suggestions will continue to be useful in future exploration of this topic.

Marilyn and Ted Helwig, for their support in countless ways, and without which I would never have been in a position to write a thesis.

Julia Pemberton, for the conversation that got the ball rolling, and the patience to help me see it through.

Table of Contents

Introduction

This thesis discusses two topics—the philosophical notion of emergence, and biology—and asks whether a particular, fundamental biological feature—life—should be deemed emergent. Emergentism is not yet at home in science, but there has been a marked increase of late in attempts to introduce it into scientific terminology. This is most common in the area of physics, where the goal is to understand the universe and its contents at the most basic level. However, it is possible to examine claims of emergence at higher levels as well, and the unique nature of life on Earth makes it a prime candidate for emergence. While the question whether life is emergent is by no means decisively settled, I will argue in this thesis that, given reasonable characterizations of life and emergence, life should not be considered emergent.

In order to probe this area properly, we need first to know what is meant in saying that something is alive. We can only follow our intuitions so far here; they act as a reasonably sufficient guide for many of the macroscopic entities with which we are already familiar, but they quickly falter when we consider microscopic, unicellular entities whose processes are superficially very different from our own. Relatively recent advances in genetics and molecular biology allow us to explore in much greater detail possible traits that may be common to all life forms. In the first chapter I use these findings to arrive at a definition of life that identifies three necessary features, and that can then be used to examine claims of emergence.

But a concrete understanding of what counts as living is only half of what is needed before the desired analysis can take place; we still need to understand what emergence means. This is not an easy job, and it is rarely explicated in the scientific literature in which it is used. Thus, it is necessary to develop a robust notion of emergence, and to distinguish this form of emergence from the weaker versions that pervade the scientific literature. To accomplish the first task, in the second chapter, I appeal to Jaegwon Kim's characterization of the received view of emergence. This provides a strong formulation that includes key tenets that demand more of proposed emergents than mere increasing complexity. To accomplish the second task (diagnosing the common, weaker form of emergence), in the third chapter I review an exemplary case to show that the weak version of emergence differs substantially from the robust, philosophical notion of emergence I apply to life.

With the requisite treatments of the topics of life and emergence completed, I move on to argue that life should not be considered emergent. There are two reasons to think this, both of which are dependent on the necessity of certain tenets in the strong formulation of emergentism that I adopt. I end by noting that possible objections to my argument may include formulations of emergentism that leave out or reject some of these tenets.

Chapter 1: Life

1.1 Introduction

What is life? This simple question is both fundamental and daunting, and we have learned that there is no easy answer. It is fundamental because, for thousands of years, humans have posed this question and offered in return views and speculations about the commonalities of life that yield insight into past cultures, their spiritual and religious beliefs, and their understanding of the natural world; and, it is daunting because life is incredibly complex. Yet at least some level of conceptual clarity regarding life is necessary if one is to be able to examine life as an emergent property. And that is, indeed, the goal of this chapter: to arrive at a characterization of life that will allow for later analysis of life in the context of emergence. To get there, in the next section I will begin by discussing how even a very basic view of life is able to discern similarities between life forms. This intuition is well supported by developments in the biological sciences. Once this connection is established, I move on, in the third section, to discussing some proposed definitions of life, followed by developing a definition of my own which distills several different scientific views of life down to three key components. I also argue in the third section that there is a basic unit of life in which these features are found. In the final section, I apply my definition to four common "borderline" cases of life to see how they stand according to the features I have selected.

1.2 Life in Common

I want to begin with a hypothetical scenario in which we must search for the criteria of life from scratch. The most immediate form of life with which we are acquainted is ourselves—that is, humans. This familiarity provides a starting point from which we may look outwards, cautiously, in search of possible connections between all life on Earth. We are, of course, already quite aware of the cycle which begins at birth, and after which we are inexorably drawn towards our death. In between those two extremes—the bounds of individual existence, individual life—we have a great many experiences, most of which will shape our character in some way, but very few of which are reflected in our bodily appearance. Apart from growth, that seems to occur as well in other humans of the same approximate age and for the same approximate duration, our physical characteristics appear mostly to be fixed at birth. Perhaps the most notable event within an individual life is the physical union undertaken with another individual life, the general purpose of which is to create the beginnings of a new individual life, to continue the cycle. Already, then, we have a very simple model of life, at least insofar as we know it from direct experience.

This four-point model of human life, consisting of birth, growth, sexual procreation, and death, is exceedingly crude, but it captures many of the similarities we find as we expand our search outwards from humans to other potential life forms. How best to expand this search is a difficult problem, but if we use as a rough guide the template generated above, we would have a considerable degree of success. This pattern, for example, would pick out almost all mammals, reptiles, birds, insects, and plant life, though a generous interpretation of sexual reproduction is required in the case of many plants, and we would run into further problems with species that reproduce asexually. Still, even in these cases we certainly want to say that these animals are alive; the problem is how to do it without tying ourselves in conceptual knots. We could loosen the restriction that reproduction be sexual, to allow for asexual reproduction, and this would seem to be progress[1]. Perhaps we could also lose the requirements of birth and death, because we know of nothing that is able to grow and reproduce that isn't also born[2] and doesn't also die or end. Should we rest, then, with this winnowed variation of the original cycle of human life, accepting as living only those entities that fit this pattern of growth and reproduction? It is clear today, with the benefit of much scientific knowledge, that we cannot, but it is equally clear that we should not dismiss the pattern's hints too quickly.

There are other patterns we have discovered and investigated, patterns which indicate, amongst other things, and to the surprise of many people, a common origin for *all* life on Earth; this is the picture painted, for example, by both evolutionary biology and molecular biology. An extensive, though not exhaustive, summary of these reasons is found in Morowitz (1992). To anticipate points of the discussion that will follow, several noteworthy patterns are described below.

- The water content of functioning living forms varies from 50% to over 95% (p. 41).

---

[1] Progress only insofar is it moves us towards what we know to be life if we step outside the human-first approach.

[2] Provided that we make the definition of born sufficiently broad, something like "comes into existence".

- There is a universal set of small organic molecules that constitute a large portion of the total mass of all cellular systems (p. 44).
- There exists a universal network of intermediate reactions such that the metabolic chart of any extant species is a subset of the universal chart (p. 49).
- There is a universal type of membrane structure used in all biological systems (p. 55).
- Every replicating cell has a genome made of deoxyribonucleic acid that stores the genetic information of the cell, which may be read out in sequences of ribonucleotides and translated into polypeptides (p. 56).
- All populations of replicating biological systems give rise to altered phenotypes that are the result of mutated genotypes (p. 57)

Perhaps the most profound question this enables us to answer is why we should think that there is a singular type of life, with characteristics in common, at all. This is, I think, something we now take for granted, but it not at all an obvious conclusion based on observation alone. It is, after all, perfectly plausible to think that life could have originated many times, each with its own evolutionary path; thus, were conditions in the past (significantly) different, the extant species we observe today could represent many different *kinds* of life. However, the patterns described above, along with still others, paint a picture of life with a single point of origin and a single, distant, common ancestor, the fundamental (and selectively beneficial) features of which were continually passed vertically through countless generations. It is the common features which have not changed at all, at least from the time of that common ancestor forwards, that we seek to enumerate and elucidate when looking for a characterization of life.

1.3 Components of Life

It is obvious from the reflections of the previous section that we share a connection with all life forms, and that this connection is best understood in the language of certain branches of microbiology, namely, molecular and cellular biology. If *what it is to be alive* is related to this connection, as I have already suggested, then it is reasonable to expect that a competent characterization of life should involve concepts and explanations from those same biological disciplines. While we may take away some hints as to what direction to investigate from the pattern of life followed by our own species (and observed in others, too), we need to find out whether there is any deeper biological explanation that will help in composing a robust notion of life.

Before investigating some existing definitions of life to see what might be gleaned, I want to introduce and adopt the terminology of Mahner and Bunge (1997) in order to clarify the discussion in general, and, in particular, what precisely is being sought in this chapter. Mahner and Bunge set forth a hierarchy of definitions, each building on and incorporating the previous ones. They begin with the general term *biosystem,* which is defined, quite rigorously, as follows (pp. 141-142):

> There are concrete systems of a kind *B* such that for every member *b* of *B,*
>
> (i) *b* is composed of chemical and biochemical subsystems, in particular water, proteins, nucleic acids, carbohydrates, and lipids;

(ii) the components of *b* are sufficiently contiguous so as to permit continual (bio)chemical interactions amongst them;

(iii) the boundary of *b* involves a flexible and semi-permeable lipid membrane (biomembrane);

(iv) *b* incorporates some of the biomolecules it synthesizes (rather than releasing them immediately to its habitat)

(v) the possible activities of *b* include the assembly, rearrangement, and dismantling of components (which allow for the self-maintenance of *b* over a certain time) as well as the capture and storing of free energy (e.g., in ATP molecules) for future consumption (metabolism);

(vi) some of the subsystems of *b* regulate most of the processes occurring in *b* in such a way that a fairly constant *milieu interieur* is maintained in the system (homeostasis, self-regulation);

(vii) one of the subsystems of *b* involved in self-regulation—its genic system—is composed of nucleic acid molecules, and its interaction with other subsystems of *b* (co)regulates the self-maintenance, as well as the development, if any, and the reproduction, if any, of *b*;

(viii) all of the control systems in *b* are interconnected by chemical signals (such as the diffusion of ions, atoms, or molecules, and propagating chemical reactions) and thus constitute a (chemical) signal network;

(ix) *b* can adjust to *some* environmental changes without jeopardizing its continued existence.

I will take up the definition of *biosystem* given above shortly, because, ultimately, it will form much of the backbone of the description of life on which I will settle. Mahner and Bunge (1997, p. 142) go on to say that "[t]he systems of kind *B* referred to [above] are called *biosystems, living systems, living things,* or *living beings* [emphasis original]," and the collection of all "living systems" are what they take to be *life*. The *elementary biosystem* is the most basic unit of life, and thus forms the base of the hierarchy; they define it as: "any biosystem such that none of its components is a biosystem" (Mahner and Bunge, 1997, p. 146). A collection of these unitary biosystems is called a *composite biosystem*, which is "any biosystem composed of (at least two elementary) biosystems" (Mahner and Bunge, 1997, p. 147). The final distinction they make represents "largest 'unit of life'", the *organism*: "An organism is a biosystem (whether elementary or composite) which is not a proper subsystem of a biosystem" (Mahner and Bunge, 1997, pp. 147-148).

Having taken these terms on board, the original motivation for this chapter can be restated. Rather than attempting to answer the very broad question "What is life?" we can focus only on the most basic unit of life, the elementary biosystem, and ask how it is realized. We can safely ignore, for the purposes of this chapter, questions about the higher-order forms of life, because as composite biosystems or organisms (of the non-elementary biosystem kind) they must be comprised of living things. It is trivial to note that these more complex forms of life will differ in many respects from elementary biosystems, but as they are themselves biosystems, I do not believe we need to explain what makes them living beyond noting that they meet all criteria outlined above.[3]

The extensive definition of biosystem already provided is an excellent starting point for assessing which biological features to include and which to exclude in a reasonable definition of life; as I have stated, it will be a refined version of Mahner and Bunge's list upon which I will settle. However, the viability of trying to list all relevant features of life has been questioned, and, I think, for good reason. Taken to its extreme, we find definitions consisting of a single quality or parameter which are supposed to explain everything, including origin (Popa, 2004). Of course it is unlikely that such simplistic definitions will suffice, but I disagree with Popa that, in general, all parametric definitions are doomed to fail. Provided that we are willing to revise the parameters we choose if given cause to do so, I can see no reason why we should not adopt a parametric definition, similar in structure to the definition of biosystem, and tinker with it according to how it performs in the field. Indeed, to foreshadow the topic of the next section, how we assess the many borderline cases of life that exist based on our definitions will tell us much about the definitions' adequacy.

Though I began the previous section with a caricatured version of how we might start to search for the criteria of life, we do not in reality need to begin with a blank slate. There is enough of a consensus in molecular biology today regarding many of the processes of life that we can state now with confidence what years ago would have

---

[3] Mahner and Bunge (p. 149) state that the "organism is, by definition, the largest living entity", which excludes the existence of "superorganisms", such as colonies or societies, that are themselves living units. Whether one agrees with this or not, it does not affect the argument of this chapter; however, I do not find their reasons for denying superorganisms to be compelling.

been controversial. (This is not to say that there is a consensus as to *which* processes are necessary to life but rather that the details of how many of these processes work and relate to each other at the molecular level is well understood.) Before returning to the definition of biosystem, I want to present several characterizations of life that will be of assistance in analyzing and pruning the definition proposed by Mahner and Bunge.

Popa (2004) lists three dominant models concerning the origin of life. What is of interest to this discussion is the observation that the models agree on the necessary features of life, but disagree as to the order in which they came into existence. The importance of these features—cellularization, replication through genetic mechanisms, and metabolism—must have been "primordial, because *none of them was dispensable* [emphasis mine]" (Popa, 2004, 11). He goes on to critique these models, saying that the question of particular importance is not which came first, "but rather what type of mechanism they were represented by in different evolutionary episodes." Maynard Smith and Szathmáry (1999, p. 4) place similar emphasis on "multiplication, variation, and heredity", for the reason that "they are necessary if a population is to evolve all the other characteristics that we associate with life." Morowitz (1992, p. 8) holds that the cellular barrier, and in particular the thermodynamic isolation it affords, "is an irreducible condition of life". His 1992 monograph is titled *Beginnings of Cellular Life*, not *Origins of Life,* because "the only life we know for certain is cellular" (Morowitz, 1992, p. 12). Schejter and Agassi (1994, p. 8) claim that the "unique characteristic of life" is "the (entropically unfavorable) permanent transfer of matter against a gradient." This is perhaps an example of the parametrically-extreme definition against which Popa cautions, but it serves to indicate the importance to life of thermodynamic considerations. Indeed there are a great many attempts to characterize life in the scientific and philosophical literature—Popa (2004) has cataloged over 100 definitions—and there are numerous examples found there that support the theme already developed by Mahner and Bunge. A few of these, found in Popa (2004, pp. 199-205), are provided below.

> Any system capable of replication and mutation is alive.
>
> The criteria of living systems are: metabolism, self-reproduction, and spatial proliferation. The more complicated kinds also have the ability to mutate and evolve.
>
> Life is defined as a system of nucleic acid and protein polymerases with a constant supply of monomer, energy and protection.
>
> In order to be recognizable life must: 1. be a non-equilibrium chemical system; 2. contain organic polymers; 3. reproduce itself; 4. metabolize by itself; 5. be segregated from the environment.
>
> We propose to define living systems as those that are: (1) composed of bounded micro-environments in thermodynamic equilibrium with their surroundings; (2) capable of transforming energy to maintain their low-

> entropy states; and (3) able to replicate structurally distinct copies of themselves from an instructional code….

We can now return to Mahner and Bunge's definition of biosystem, to see how it compares to the various definitions of life given above. Point (iii) is certainly the easiest to examine, since it clearly picks out the idea of cellularization. Points (iv) through (vi) can be grouped together under the heading of metabolism, with the additional note that the type of metabolism suggested there requires the sort of semipermeable membrane dictated by point (iii). Points (vii) and (ix) specify the requirements for replication and adaptation (the latter requires heritability and variability), and will be useful in the discussion of examples of non-reproducing life forms found in the next section. Point (ii) seems not to pertain to particular characteristic life—though it is certainly necessary—so much as to assist in distinguishing one biosystem from the next; it is reasonable, I think, to consider (ii) as implicitly included in the definitions provided above, having been spelled out explicitly by Mahner and Bunge for the sake of formality. This may be said of point (viii) as well, which describes the necessity of a chemical signal network within the biosystem. This leaves points (i), which seems to be simply a contingent fact based, in large in part, on Earth's environment during biogenesis, as well as the actual historical (rather than chemically or physically possible) path travelled by evolution. There is one additional point which needs to be added to the original definition because, as already noted, we are interested primarily in elementary biosystems. Therefore, I shall add point (x), which states that for a biosystem to be elementary, none of its components can be a biosystem.

The recurring components throughout the definitions of life can now be identified. They are cellularization, metabolism, and heredity[4], and they form the basis of the definition of life that I will use in this thesis. Explicitly, these features were selected for their prevalence in the literature on the origin and definitions of life, combined with their observed universality in all life examined to date. Each can be broken down further—exemplified both in some of the points listed by Mahner and Bunge above, and by a further clarification to the notion of heredity that will be presented in the next section—but as fundamental features of life, so far as we know, they must be present.

Three things remain to be accomplished in this chapter: first, to specify briefly the biological role of each component; second, to suggest and discuss the basic unit of life in which these three components are found; and third, to assess the field competence of this definition by using it to examine several problematic candidates for life.

Cellularization

The possession of a means of separating an entity from its environment is undeniably a necessary component of life (Popa, 2004; Morowitz, 1992). Some entities, such as viruses, are protected by a very simple barrier—often only a protein coating—the sole function of which is to protect the virus's genetic core from hydrolysis. More complex entities, especially those with active metabolisms, require membranes that allow for selective transportation to occur across them. These membranes are spherically

---

[4] As will become clear, I follow Maynard Smith in insisting that heredity be accompanied by variation. As a shorthand I will simply use heredity, but all further instances of heredity should be taken to include variation as well.

shaped amphipatic bilayers, consisting of two fatty acids arranged such that their hydrophobic ends are buried in the center of the bilayer, while the hydrophilic ends face outwards into the aqueous solutions that make up the interior of the cell and the exterior surrounding the cell. Some of the benefits of cellularization are listed below.

- thermodynamic protection, resulting from “increased ability to withstand hydrolysis” leads to increase in the entity’s half life (Popa, 2004, p. 53)
- creation of directional flow of energy by confining catalyst activity to primary vesicles (Popa, 2004).
- creation and maintenance of “sharp (i.e., energy rich) gradients, used by life as energy currencies. The osmotic arrest of large molecules...also allowed the confinement of large [nucleic acids] which could serve as genetic materials” (Popa, 2004, p. 55)
- possible impetus for creation of the genome, by confining genes together which benefit from all genetic material being copied (Maynard Smith and Szathmáry, 1999).

Metabolism

Even the most basic empirical analysis of life must recognize immediately the importance of an intake-process-excrete cycle. On a macroscopic scale, it is manifested in numerous ways (e.g. eating food). At the molecular level, metabolism is a sequence (pathway) of chemical reactions within a biosystem that are, notably, assisted by numerous enzymes. Enzymes are (typically) proteins that increase the speed and efficiency of the reactions they catalyze, enabling reactions which are thermodynamically unfavorable to occur. The products of metabolism vary greatly depending on where the reaction sits in a metabolic pathway; intermediate products are often used a substrates in the next reaction in the chain, while end products have specific roles within the cell, such as "an amino acid that can be incorporated into a polypeptide, or a sugar that can be consumed for its energy content" (Karp, 2005, p. 108). In short, metabolism provides biosystems with (among other things) a means to grow and self-regulate, two features which are important both macroscopically and microscopically.

Heredity

It is not enough to insist *only* that life must replicate (or multiply). Some kinds of replication simply would not work for life—for example, replication exhibited by crystals or fire (see next section), or by some other process, such as from a mold. Importantly, all life forms on Earth exhibit heredity—that is, they pass along, through various means, information ("blueprints") for, or that contribute to, the construction the next generation of biosystem. Over time, however, this information can slowly become changed by the accrual of small, statistically inescapable genetic mutations (variations) (Morowitz, 1992). These mutations may or may not have an effect on the biosystem's ability to reproduce, but if they do, those mutations which produce a favorable reproductive effect will, in time, become dominant. This is the familiar concept of *natural selection*. Thus, both heredity and and variability are required if natural selection is to take place, as we observe it to.

These features of life which I have identified as necessary can be stated explicitly as three postulates.

> Postulate 1: living things must have heredity with variation (or, equivalently, be subject to natural selection);
>
> Postulate 2: living things must have a metabolism;
>
> Postulate 3: living things must have a selectively permeable barrier separating them from their environment;

Crucially (and perhaps obviously), by stipulating the necessity of each of the above features it follows that they must always be found together for the system to be considered living. The next section will examine some cases where one or more features is missing.

Empirical observation supports the conclusion that the most fundamental unit in which these features are found is the cell (Hurst, 1937; Eigen, 1992; Morowitz, 1992, 2002; Popa, 2004, Mahner and Bunge, 1997). This can be restated in the following way (Mahner and Bunge, 1997, p. 146):

> All elementary biosystems are cells.

The converse, however, is not necessarily true. A classic example is the eukaryotic cell, which contains several organelles (mitochondria, chloroplasts) that appear to be former bacteria that have been captured and enslaved over the course of evolution. Another example, given by Mahner and Bunge (1997, p. 147), comes from the single-celled ciliates, which contain single-celled "algal symbionts, such as zoochlorellae. Clearly, these algae are not only biosystems but also components...of a ciliate cell. Thus, when containing zoochlorellae…[a cell] is not an elementary biosystem." Morowitz (1992, p. 39) sees the view that cells are the basic units of life as "so much a part of an overview that it is rarely necessary to make explicit." He also describes in detail the mollicutes, "a class of wall-less prokaryotes that are judged to contain the smallest cells by visible microscopy, electron microscopy, and filtration through small, pore-sized barriers" (p. 59). These deserve further consideration, since they point to the likely existence of a lower bound on size and simplicity which supports the notion that cells are in fact the elementary unit of life.

The first thing to note about some Mycoplasmatales is their diameter, which can be "as small as the order of 0.3 microns" (Morowitz, 1992, p. 59). He continues,

> If we consider the hardware necessary to perform [various tasks including metabolism, membrane construction, protein synthesis, and replication], the need for enough atoms clearly imposes a lower size limit imposed by atomicity. For example, consider a hypothetical cell half the diameter of the small mycoplasma. The radius would be 0.075 *μ*, and since the membrane is about 0.005 *μ* in thickness, the inner sphere would have a

> radius of 0.07 μ. The nonaqueous portion of the interior would contain 4 x 10^7 atoms. Since the average macromolecule contains about 10^4 atoms, the cell contains only four thousand total macromolecules of all types. Alternatively, this number of atoms could make up a small genome, ten ribosomes, and the order of one thousand protein molecules. Clearly, this is pushing an absolute limit for the number of functions that must be performed by a cell.

Its small size also limits its metabolic complexity, because much of its genome is taken up by regions which code for the machinery of replication. Morowitz estimates the number of proteins coded for by the DNA of mycoplasma at about 550 (329,000 amino acids at approximately 600 amino acids per protein); this further "defines a level of functional simplicity at which cells can operate, since proteins, with some notable exceptions tend to have single enzymatic functions and may have one or more control factors" (Morowitz, 1992, p. 62). He concludes his discussion of mycoplasma by saying that "it is apparent that these forms fit all the criteria for life...and have reduced the necessary hardware for survival to something near a minimum" (Morowitz, 1992, pp. 65-66).

Though the case for the chosen criteria is well supported and its conclusion, that all elementary biosystems are cells, is well accepted, it is important to examine some stand-out cases in which the question of life is open. This is the subject of the next section.

1.4 Examples

Taking stock, we now have a working template for life, which consists of three key features (cellularization, metabolism, and heredity), along with an idea of what the most basic unit is in which all three features are found (the cell). And we do not say that *only* cells are alive but, rather, that anything comprised of, at least in part, (functioning) cells is alive, too; thus we have both living single- and multi-cellular organisms. In essence, by emphasizing the importance of the cell as the basic unit of life, we have allowed for the label of living to be applied to all organisms which we would want to claim as living *a priori*, because they are all, loosely speaking, large collections of cells working together. But, we have also enabled the application of the term *living* to microscopic organisms, with which our cells, as individual units of life, have much in common. We are all a part of one enormously extended family, and it is for this reason that we have life in common.

It is no good to have a definition of life that is useless in practice. There are many organisms which sit on the fence between life and non-life, with their status, often as not, being determined by the specifics of the definition of life one adopts. We have seen this in practice earlier, where I examined life by considering the human life cycle as a starting point; though such a definition captured many macroscopic (multi-cellular) extant species which we would certainly want to think of as living (mammals, reptiles, birds, fish, and so on), it was also too strict. By moving towards a more fundamental definition, the number of species included within its definitional boundary increased dramatically. Now, we need to explore its proposed borders. Thus, in this section I want to examine some borderline cases by making use of the definition I have developed so far. Each of the examples were selected both for their prominence—they are often trotted out in discussions of life—and for the particular component of the definition they emphasize. Thus, these considerations should be taken as illustrative of what should be included as living under the definition I have proposed, and what should be excluded.

Fire

The first case to consider is fairly simple: fire. Most people, and especially those who have not reflected on what it means to be alive, do not think fire is living. Why, then, should it even be considered? The primary reason is that the example of fire is, I think, the easiest way to demonstrate the importance of the *joint* possession of the features of life identified in the particular definition I have chosen. Fire clearly has a metabolism: it receives fuel which, after undergoing chemical reactions, is changed into products (mostly in the form of water and carbon dioxide) that then leave the system. This is metabolism as its most basic level, and it is in this way that fire can grow, provided it is given a constant supply of fuel. If the fuel supply is turned off, the fire quickly dies out. This can be compared to the necessity of food and water for animals, or to the availability of many key molecules in metabolic pathways (such as glycogen and ATP). In the absence of a stable supply of these products, the metabolic pathways shut down and life is eventually extinguished.

Fire can also replicate, a fact made trivially evident by observance of (the flames of) lighters and matches spawning “daughter fires”; furthermore, these daughter fires even display variation, both between each other and alone, through time. Interestingly, if we adopted the simple human cycle as a definition of life, it would seem that fire must be considered alive as well. So how does fire stack up against Postulates 1-3? Most obviously, fire lacks both heredity and cellularization, both of which are necessary for fire to be deemed alive. While fire is able to replicate, “the characteristics of a fire depend only on the supply of fuel and oxygen at the time, and not on whether the fire was lit by a match or cigarette lighter. Lacking heredity, fire does not evolve, and so lacks the adaptive complexity that only natural selection can confer” (Maynard Smith, 1999, p.5). Nor does fire have any concrete physical structure associated with it; it is simply a physical space in which the fueled chemical reactions are taking place, with no boundary (but, rather, the characteristics of the fuel source) to confine the reaction space. We can conclude that, in lacking two crucial features, fire cannot be alive.

Viruses

The next case to be examined, viruses, is, arguably, the most frequently debated, because it is often found on the fence between living and non-living. It is a textbook example of a system whose status is determined by the subtleties of the definition of life one chooses. A description of viruses is given by Karp (2005, p. 21):

> Outside of a living cell, the virus exists as a particle, or virion, which is little more than a macromolecular package. The virion contains a small

> amount of genetic material that, depending on the virus, can be single-stranded or double-stranded, RNA or DNA. Remarkably, some viruses have as few as three or four different genes, but others may have as many as several hundred. The fewer the genes, the more the virus relies on enzymes and other proteins encoded by its host cell's genes.

It seems that the controversy surrounding the status of viruses is primarily the result of their being acted on by natural selection. Viruses are infamously known for their ability to adapt at a rapid pace, making attempts to contain or destroy them exceedingly difficult. Their small genomes lead to high mutation rates in the copies produced, so that mutations that exhibit immunity to anti-viral agents quickly become selected for. When this fact is combined with their high speed of replication, viruses become perfect case-studies in natural selection. On some accounts of life, the ability to be acted on by natural selection is sufficient for the entity to qualify as living. Because viruses mimic so closely the patterns of replication we observe in all life forms, it is easy to understand the confusion.

However, what is notably absent is an autonomous metabolism. They are mere replication-machines, ultimately dependent on the host cells they invade for their replication to take place, though some (for example, plus-strand RNA viruses) are more dependent than others (minus-strand RNA viruses, retroviruses). In particular, the molecules utilized for replication are located within the confines of the host cell's membrane, and it is the host cell's own replication machinery (with the addition of an introduced replication enzyme for minus-strand viruses) which carries out the replication. By contrast, the amino acids used by cells are often obtained through various internal metabolic reactions—and this is precisely what is hijacked by the virus. Thus, because viruses lack their own metabolism, we can say that while the cell-virus system is alive, the virus itself is not. This is echoed by Karp (2005, p. 22): “virions are macromolecular aggregates, inanimate particles that, *by themselves*, are unable to reproduce, metabolize, or carry on any of the other activities associated with life. For this reason, viruses are not considered to be organisms and are not described as being alive. [emphasis mine]”

Sterility and Dysfunction

The third example I want to consider presents an interesting challenge to the definition of life offered in the previous section. Recall that this postulate requires the joint possession of the features identified by postulates 1-3 in order for something to be deemed living. However, there are numerous examples of entities that, often as a result of mutation, do not (strictly speaking) have one of the required features. It is evident that the lack of a barrier is almost certainly a fatal mutation, and need not be considered. So too, it seems, is the lack of a metabolism. It is exceedingly unlikely that such large mutations would occur, at any rate; much more likely than wholesale absence is small-scale mutation resulting in dysfunction. It is this sort of mutation that we must consider when we examine particular (or *token*) biosystems with faulty membranes, metabolisms, or means of reproduction. These token biosystems should be differentiated from the biosystem’s *type*, which generally includes working versions of these subsystems. We can partially alleviate this concern by stipulating that "some of the properties of biosystems are dispositions that may actualize under favorable circumstances" (Mahner and Bunge, 1997, p. 143). Systematic dysfunction, then, merely renders the circumstances for actuation of the particular feature *un*favorable. Thus, infertile members of species that otherwise exhibit the feature of heredity are clearly living, despite the fact that, as instantiations of a biosystem, they are unable to reproduce; the same goes for tokens which have faulty metabolisms, or structural deformities.

However, there is a related but distinct problem in the form of subsets of a population which, through evolution, have lost one or more of the key features. Sterile castes of insects fit this description, as do erythrocytes, which do not carry the genetic material necessary for replication or protein synthesis. In the case of sterile castes of insects, it is not a problematic mutation that has led to their inability to reproduce; their sterility is, for their functional purposes, an evolutionary feature of the subset of the species of which they are members. The crucial point that allows this case to be decided lies in the composition of the notion of heredity. Until now I have considered heredity to be tied intricately with replication, hence the current problem. However, what is often missed due, I think, to its implicit nature is the fact that biosystems are also products of heredity. So, insofar as we are concerned here, heredity is a two-part concept; being on the receiving end of heritable transmission simply takes a back seat to replication with heredity most of the time. In most instances it is not necessary to spell out the former requirement because it is always the case, but frequent omission may leave it seeming less important than it really is. In fact, it is the *true* requirement for heredity, and it is only because it so often results in successive generations passing along their genetic information in the same way that reproduction-with-heredity is identified with life at all. All individuals in sterile castes were passed genetic information in the same way as their fertile neighbors. The difference between them lies in what that information specified. We do not need to insist that each member be able to replicate—we have seen already that particular mutations make this impossible—but instead that each individual be the result of reproduction in which genetic information was passed along, regardless of the reproductive capabilities it specifies.

It is worth mentioning, if only briefly, that the above requirement of being a product of heritable transmission may appear to impose conditions on life that depend not only on the properties of a biosystem at a given time, but also on its history. We can get around this by noting that we only know biosystems that are a product of evolution; we never see complex entities that exhibit the requisite characteristics of life arising *de novo*. Thus, in observing such entities, we can safely assume that they are indeed products of heritable transmission.

Artificial Life

The final challenge I will examine is artificial life (AL). It is most often conceived today as computer-based (artificial intelligence, or AI), but it need not be. Proponents of *strong* AL have as their goal the synthesis of “life-forms alternative to the ‘carbon-chain life’ as known to biology ‘by attempting to capture the behavioral essence of the constituent components of a living system, and endowing a collection of artificial components with similar behavioral repertoires…'" (Mahner and Bunge, 1997, pp. 149-150). In their analysis of the strong AL program, Mahner and Bunge, quoting Rasmussen, note the irrelevance to its supporters of the composition of the candidate systems: "...the ontological status of a living process is *independent* of the hardware that carries it" (Mahner and Bunge, 1997, p. 150). Their critique (pp. 150-151) of the strong AL program is based on the fact that

> the biosystems on our planet are the only living things we know. Hence, whatever artificial system is presumed to exhibit some property or properties of (genuine) biosystems, it can be compared only to the biosystems known to biologists. Thus, AL technologists cannot show that they have succeeded in constructing genuinely *alternative* life forms because such artificial systems, if exhibiting...properties others [sic] than those known from life on Earth, could be said to be alive only *by definition*: they would not be *found* to be alive but *declared* to be alive. Hence...it is not just 'extremely difficult' to distinguish universal properties of biosystems (as they could be) from those common to life on our planet: it is *de facto* impossible. [emphasis original]

A full treatment of this problem is too lengthy and tangential for my purpose here[5], but I agree with the spirit of their critique. It is motivated by the picture of life currently

---

[5] The main reason for it being tangential is that it involves discussion of how to treat man-made objects exhibiting high-level characteristics. It is virtually impossible for complex life forms, such as a proposed intelligent computer, to spring into existence without passing first through some evolutionary pathway

available to us which, given the way life has developed on Earth, *of course* exhibits particular commonalities. It is only by observing these commonalities and generalizing from them that we are able to suggest definitions for life in the first place. But if we were to observe not a fundamental aspect of life but an advanced feature, such as apparent consciousness, we would be justified in at least suspecting the presence of life, regardless of the details of the entity itself. In other words, the definition of biosystem given by Mahner and Bunge only dictates a minimum requirement (which is couched, no less, in the language of life particular to Earth); we can easily imagine other sufficient (but not necessary) features which could also indicate life. In their conclusion they assent to the idea that extraterrestrial life may exist that would very likely differ drastically from Earth's life forms, though they remain unwilling to extend the same consideration to terrestrial artificial life, calling it as a project a "waste of energy, time, and money" (Mahner and Bunge, 1997, p. 153). However, given our current level of technology (and the rapid pace of its increase), it seems unfair to rule out entirely the future creation of some entity that perhaps exhibits some higher-level features of life while lacking those already identified in this chapter; at that time, we would likely have to revise our definitions of life to include, explicitly, higher-level features that we currently assume imply the fundamental features we have deemed necessary for life. But as this is currently only the domain of science fiction, speculation about the status of hypothetical entities seems unlikely to contribute anything of significance to talk about the emergence of life.

(even indirectly, by considering their creator). It is not clear, therefore, how such cases would be handled were they to be found.

Chapter 2: Emergence

2.1 Introduction

Because this thesis is concerned with addressing emergence as it may be applied to life, it is therefore necessary not only to understand life—covered in the first chapter—but also the philosophical position of emergentism. The next section will begin this task by stating and examining the central tenets of emergentism. However, two of these tenets will, because of their importance to this thesis, require a greater degree of scrutiny. Thus, the third section will cover the topic of reduction by looking at two different models, while the fourth section will deal with the causal efficacy of emergent properties. My hope is that by examining emergence with respect to its particular tenets, rather than as a wholesale doctrine, it will be easy to identify and incorporate points of agreement in the next chapter, while allowing for diagnoses of particular problems where they arise.

2.2 Emergentism

It is quite common to find in modern philosophical discussions of emergence some acknowledgment of its relatively recent rise in popularity. Talk of emergence is not limited to philosophical discourse, however; it is used in many different contexts within the natural and special sciences to denote a variety of ideas, so that its meaning from one context to the next can differ significantly. In order to explore possible connections between emergence and life, it will first be necessary to acquire an understanding of emergence that respects its historical underpinnings while remedying, if and where possible, theoretical missteps both original and more recent. My goal is neither to provide a complete review of the literature on emergence nor to develop yet another specific brand of emergentism; I will focus instead on presenting what is often taken to be the generally received view of emergentism, along with some criticisms, where relevant.

The best place to begin, I think, is by stating in very rudimentary form the general concept of emergence. An excellent formulation of this sort is offered by Jaegwon Kim (2006, p. 548), and it introduces and relates several key features of emergentism that will discussed at length in the following sections:

> the intuitive idea of an emergent property stems from the thought that a purely physical system, composed exclusively of bits of matter, when it reaches a certain degree of complexity in its structural organization, can begin to exhibit genuinely novel properties not possessed by its simpler constituents.

This “intuitive” formulation of emergentism gives insight into its proponents’ view of how the physical world is organized. The degrees of organizational complexity roughly correspond to “levels” so that the world is divided up into a hierarchy of these levels. It is convenient (though questionable, as we will see) to think of these levels in terms of sciences that study them. Thus, at the base of the hierarchy is physics, which studies the fundamental constituents of matter and the ways in which they interact with each other. Physics, like other sciences, is characterized by a set of laws which are thought to describe these fundamental interactions. Setting aside the question whether physics itself can be divided further, the next tier in the hierarchy is typically taken to be chemistry[6], though sometimes the base level is described as a hybrid, “physico-chemical”. Next up we find biology, after which the structure of the hierarchy becomes much less obvious. The relevant point is the supposition that at each tier, and presumably for sub-tiers if they exist, properties emerge that somehow go beyond what we might expect (hence “novel”) based on our knowledge of the laws of the level below.

In order to examine this idea fully, however, we need to find a more complete description of just what emergentism claims. For this, too, we can turn to Kim (1999, pp. 20-22) who summarizes the main tenets of the “received” view of emergentism.[7]

---

[6] More accurately, it should be deemed chemical. However, to remain consistent while avoiding confusion by calling the base tier “physical”, which also clearly applies to higher tiers as well, I will simply continue to use the name of the relevant scientific field to identify the levels.

[7] Though it is never stated explicitly, tenets 3-5 should be seen as necessary features of emergentism. And, insofar as there are complex higher-level entities and properties, they should be seen (necessarily) as arising according to tenets 1 and 2.

(1) *Emergence of complex higher-level entities:* Systems with a higher-level of complexity emerge from the coming together of lower-level entities in new structural configurations (the new “relatedness” of these entities).

(2) *Emergence of higher-level properties*: All properties of higher-level entities arise out of the properties and relations that characterize their constituent parts. some properties of these higher, complex systems are “emergent”, and the rest merely “resultant”.

(3) *The unpredictability of emergent properties:* Emergent properties are not predictable from exhaustive information concerning their “basal” conditions”. In contrast, resultant properties are predictable from lower-level information.

(4) *The unexplainability/irreducibility of emergent properties*: Emergent properties, unlike those that are merely resultant, are neither explainable nor reducible in terms of their basal conditions.

(5) *The causal efficacy of the emergents*: Emergent properties have causal powers of their own—novel causal powers irreducible to the causal powers of their basal constituents.

In addition to clarifying the emergentist doctrine, Kim's tenets also introduce several new ideas. We have seen points (1) and (2) in the intuitive formulation, but (3)-(5) add the requirements of unpredictability, irreducibility, and causal efficacy. Points (4) and (5) will be given their own sections, so for now we can focus on addressing points (1)-(3).

The idea that complex higher-level entities emerge from certain configurations of lower level entities is one of the oldest ideas in emergentism. There are many examples of this sort of emergence, from Broad's (1925) own atoms-to-compound-molecules, to, perhaps, organisms from cells, and so on. As Kim (1999, p. 20) notes, this idea is hardly unique to emergentism, and it is not until we consider the properties of these entities that we move into the realm of emergence. An important distinction must be made when we consider such properties. Emergentists distinguish between two kinds: resultant properties and emergent properties. An example, adapted from Kim (1999, pp. 25-26), involving mass can be used to gain an idea of what is being said in this distinction. Consider an atom of iron. It has certain properties, one of which is its mass. A 1 kilogram hunk of iron has a great many iron atoms in it, such that the overall mass is 1 kilogram. Though none of the individual atoms have the property of having a mass of 1 kilogram, the hunk itself does, and as such it is able to exert particular causal powers, like acting as a brutish paper weight. But despite the fact that none of its constituent parts (atoms) share the property of having a mass of 1 kilogram, "emergentists would not consider mass an emergent property; they would say that the mass of an object is a resultant property, a property that is merely 'additive or subtractive'."

The difference between resultant and emergent properties is often characterized by the "novelty" of the latter properties:

> The concept of emergence was dealt with (to go no further back) by J.S. Mill…The word 'emergent', as contrasted with 'resultant', was suggested by G.H. Lewes…Both adduce examples from chemistry and from physiology; both deal with properties; both distinguish those properties (a) which are additive and subtractive only, and predictable, from those (b) which are new and unpredictable. (Morgan, quoted from Kim, 1999, p. 5)

Because we are interested in the properties picked out by (b) in the above quotation, we must explore what is meant by "new and unpredictable". Kim suggests that the unpredictability is the source of the novelty, so long as we are concerned only with the right kind of predictability. He identifies two types of predictability—*inductive* and *theoretical*—and insists that it is the latter which emergentists deny. Of the former, he says

> Even emergent properties are inductively predictable: Having observed that an emergent property, E, emerged whenever any system instantiated a microstructural property M, we may predict that this particular system will instantiate E at t, given our knowledge or belief that it will instantiate, M, at t [endnote excluded]. More generally, on the basis of such empirical data we may have a well-confirmed "emergence law"

> to the effect that whenever a system instantiates basal condition M it instantiates an emergent E.

The notion that emergent properties are inductively predictable is something with which everyone is familiar; if we take the example of consciousness as an emergent property of a certain biological microstructural configuration (the brain), we assume that because all humans have brains they must have some form of consciousness. This type of prediction is closely tied to the notion of *supervenience,* which says that

> If property M emerges from properties $N_1,\ldots,N_n$, then M supervenes on $N_1,\ldots,N_n$. That is to say, systems that are alike in respect of basal conditions, $N_1,\ldots,N_n$ must be alike in respect of their emergent properties. (Kim, 2006, p. 550).

Supervenience is often considered an integral component of emergence, as it relates in a deterministic way basal properties and emergent properties (Kim, 2006). But could we predict, for example, consciousness in the absence of experience, by examination of the physical constituents (and their arrangement) alone? It seems exceedingly unlikely, for we would lack even the theoretical concepts necessary to generate such a prediction. We can strive to *replicate* something like consciousness based on the inductive prediction that certain complex systems of the kind we have already observed will exhibit consciousness—this is the goal of the strong AI programme—but we cannot, if consciousness is indeed emergent, hope to design novel instantiations.

2.3 Reduction

The irreducibility of emergent properties is a key tenet of emergentism because it grants special status to each level in the hierarchy. The sciences that describe the laws and explain the phenomena of these levels are not simply convenient macroscopic descriptions that reduce down to the basal domain of physics on close inspection. By contrast, the goal of reductivism is to provide explanations consisting *only* of basal laws and the constituents they govern.

I will cover two kinds of reductionism in this section. The first approach is often called Nagelian or classical (theory) reduction, so named for its most influential contributor Ernest Nagel, though I will include Fodor's (1974) characterization here. The second approach, developed by Kim (1999; 2006) is a more recent and unique formulation that we can call the functional model of reduction. Ultimately I will proceed by using Kim's method of reduction, both for its relative simplicity and its self-contained nature; it is well beyond the scope of this work to synthesize the many responses to Nagel's (and even Fodor's) original position before applying the result to biology. Instead, by using Kim's proposal for reduction, a comprehensive recounting of it can be given here, and it can readily be applied to the problem motivating this entire thesis. However, it is still worthwhile to consider the classical view, both for completeness and to offer a contrast to Kim's version.

Classical Reduction

The basic idea behind classical reduction is that higher-level theories can be reduced to lower-level (more basic) theories by making use of *bridge laws*. They act in the following way, as illustrated by Fodor (1974, p. 98). We start with a given law of a particular special science[8] S and denote the relation as

(1) $S_1x \rightarrow S_2x$

which Fodor instructs us to read as “all S_1 situations bring about S_2 situations”. In order to reduce this special science law to physics, we require a means of relating both S_1x and S_2x to their respective situations in physics. This is particularly necessary (and almost always the case) when terms in science to be reduced do not occur in physics. We can write these relations

(2a) $S_1x \leftrightarrows P_2x$

(2b) $S_2x \leftrightarrows P_2x$

(3) $P_1x \rightarrow P_2x$

---

[8] It is unclear what exactly counts as a “special science”. Here I will take it to mean any science that is not a part of physics proper (i.e., chemistry, biology, and so on). Whether Fodor intended the other natural sciences to be included is not explicitly clear, but his insistence on the transitivity of bridge laws seems to clear the way for counting as a special science any *non*-physics science.

Here P_1 and P_2 are predicates in physics, while S_1 and S_2 are predicates of the special science. Unlike (1) and (3), the bridge laws contain predicates of both physics and the special science; thus, "the reduction of a science requires that any formula which appears as the antecedent or consequent of one of its proper laws must appear as the reduced formula in some bridge law or other" (Fodor, 1974, p. 98). Fodor suggests that the bridge laws express event identities, so that (2a), for example, reads as roughly "every event which consists of x's satisfying S_1 is identical to some event which consists of x's satisfying P_1 and vice versa" (Fodor, 1974, p. 100). The doctrine entailed by this reading of bridge laws he calls *token physicalism*, which is "simply the claim that all the events that the sciences talk about are physical events" (Fodor, 1974, p. 100). To get to reductivism, we must add to token physicalism the "assumption that there are natural kind predicates in an ideally completed physics which correspond to each natural kind predicate in an ideally completed special science" (Fodor, 1974, p. 100). Fodor encourages us to accept token physicalism but deny reductivism, on the basis that it seems unlikely that there is the correct correspondence between the natural kinds of the respective sciences. Nonetheless, he offers the following summary:

> reductivism entails the generality of physics in at least the sense that any event which falls within the universe of discourse of a special science will also fall within the universe of discourse of physics. Moreover, any prediction which follows from the laws of a special science and a statement of initial conditions will also follow from a theory which consists of physics and the bridge laws, together with the

> statement of initial conditions. Finally, since ‘reduces to’ is supposed to be an asymmetric relation, it will also turn out that physics is *the* basic science; that is, if reductivism is true, physics is the only science that is general in the sense just specified. (p. 101).

There are problems with classical reduction, and more recent work has been focused on addressing them. Some corrected accounts are covered briefly in Batterman (2004), but, having only skimmed the surface of the vast literature on this type of reduction, I will set it aside in favor of the novel approach of Kim. At the end of this section I will offer some insight into why Kim’s is the preferred model for application to the topic of this thesis.

Functional Model of Reduction

Kim’s (1998; 1999) model is motivated by three crucial questions to which he feels his model, unlike the classical model, is able to provide satisfactory answers.

(1) *Explanatory*: why does this system exhibit E at t?
(2) *Predictive*: Will this system exhibit E at time t? Can we predict this from knowledge of what goes on in the base domain?
(3) *Ontological*: In what sense is the functional model a model of reduction? What does it reduce, and how does it do it?

The model itself veers away from the classical bridge laws, instead reducing a property to its reduction base by means of a three-step process. The first step is to functionalize the candidate property; once this is accomplished, we search for the realizers of this property in the basal level; finally, having identified a realizer (or several realizers), we find a theory at the basal level that explains how these realizers cause the property in question. Upon achieving the goal of the third step, the reduction is complete. With so much time and effort invested into the classical model since its inception, one may rightly be wary of Kim's seemingly simple approach; however, Kim's example, along with further explication of the steps, helps to clarify exactly how and why this unique approach to reduction works.

As in classical reduction, Kim's model is ultimately a step-wise process. The reduction base may not initially be the most general base possible; further reduction is necessary following the same steps in order to reach the level of fundamental physics. The reduction base, **B**, consists of the "basal conditions for our emergent properties", from which property *E* is proposed to have emerged. The goal is to reduce *E* to **B**, though, as noted, further reduction may be desired. The first step required for this reduction is the functionalization of *E*:

> Step 1: E must be functionalized - that is, E must be construed, or reconstructed, as a property defined by its causal/nomic relations to other properties, specifically properties in the reduction base **B**. (Kim, 1999, p. 10).

E is functionally defined over **B** as:

> Having E $=_{def}$ Having some property P in **B** such that (i) C1,…,Cn cause P to be instantiated, and (ii) P causes F1,…, Fm to be instantiated. (We allow either (i) or (ii) to be empty.)

Any property P in **B** that meets (i) and (ii) is a "'realizer' or 'implementer' of E" (Kim, 1999, p. 11). In functionalizing E, we are "establishing a conceptual/definitional connection for E and the selected causal role", and thus we should include empirical knowledge about E and its causal/nomic relations. Having functionalized E, we can move to the next step:

> Step 2: Find realizers of E in **B**. If the reduction, or reductive explanation, of a particular instance of E in a given system is wanted, find the particular realizing property P in virtue of which E is instantiated on this occasion in the system; similarly, for classes of systems belonging to the same species or structure types. (Kim, 1999, p. 11).

Here the challenge is empirical (scientific). Kim uses the example of a gene to illustrate this point: to reduce the property of being a gene (genetics) to a property of DNA (molecular biology), we must first functionalize the gene. The causal function of a gene can be roughly construed as the transmission of heritable information to offspring.

However, in order to find the realizer of this property, much scientific legwork needed to be done. DNA happens to be the modern realizer, although this may not always have been true. Fortunately, Kim's model can accommodate multiple realizers, a point which will facilitate its application to life in the next chapter. The final step in Kim's model is to identify a theory that explains how E works. Specifically,

> Step 3: Find a theory (at the level of **B**) that explains how realizers of E perform the causal task that is constitutive of E (i.e., the causal role specified in Step 1). Such a theory may also explain other significant causal/nomic relations in which E plays a role.

If we are seeking a reduction of the gene to the level of molecular biology, we need to find a theory at the level of molecular biology that explains how DNA is able to accomplish its causal function. Typically, as Kim notes, "ascertaining realizers of E will almost certainly involve theories about causal/nomic interrelations among lower-level properties in the base domain" (Kim, 1999, p. 12). Once we have identified the realizer and are in possession of a theory at the basal level that explains the causal function of the realizer, we have on Kim's account successfully reduced the target property to its reduction base.

Returning to Kim's motivating questions, we have seen already how to answer the first. The system has a property with a particular causal role, and for which there is a specific realizer and an accompanying theory as to how that realizer works. To answer the second question, we need only look for the realizers of E at the basal level.

Recalling the example of the gene, once the gene had been functionalized, the scientific problem could be recast at the basal level. The search for the particular molecular realizer of the causal function of the gene could be conducted “solely on the basis of knowledge of the causal/nomic relations obtaining” at the molecular level (Kim, 1999, p. 14). And once DNA was found, we were able to predict, based on its presence or absence, which systems would have heritable transmission (though the details of heredity, such as gene expression, are another matter).

The final, ontological question provokes a two-fold answer from Kim. First, he suggests that because each instance of E is directly tied to a realizer of E, there is nothing above and beyond the instances of the realizer. Because there are potentially many realizers of E—such as DNA, RNA, and so on—we can identify and group together all instances which possess the same causal powers. Kim calls this the *causal inheritance principle*, and formally it reads:

> If a functional property E is instantiated on a given occasion in virtue of one of its realizers, Q, being instantiated, then the causal powers of this instance of E are identical with the causal powers of this instance of Q. (Kim, 1999, p. 16).

If one grants this, then the E- and Q-instances are effectively identical, and thus all E-instances reduce to Q-instances.

The second part of Kim's answer to the ontological question concerns the reduction of the property E itself, especially in light of there possibly being multiple realizers of E. He lists three approaches:

> First, one may choose to defend E as a legitimate higher-level property irreducible to its realizers, the Q's….
>
> Second, one may choose to identify E with the disjunction of its realizers…. Notice, though, that this identity is not necessary—it does not hold in every possible world—since whether or not a property realizes E depends on the laws that prevail at a given world.
>
> Third, we may give up E as a genuine property and only recognize the expression E or the concept E. As it turns out, many different properties are picked out by the concept E, depending on the circumstances—the kind of structures involved and the nomological nature of the world under construction.

The first approach is essentially that of emergentists since the irreducibility of emergent properties to some basal level is a key tenet of emergentism. A longer passage makes clear Kim's stance with respect to the second and third approaches; it will also factor significantly into the next chapter.

> [I]f the “multiplicity” or “diversity” of realizers means anything, it must mean that these realizers are causally and nomologically diverse. Unless two realizers of E show significant causal/nomological diversity, there is no clear reason why we should count them as two, not one. It follows then that multiply realizable properties are ipso facto causally and nomologically heterogeneous. This is especially obvious when one reflects on the causal inheritance principle. All this points to the inescapable conclusion that E, because of its causal/nomic heterogeneity, is unfit to figure in laws, and is thereby *disqualified as a useful scientific property* [emphasis mine]. On this approach, then, *one could protect E, but not as a property with a role in scientific laws and explanations* [emphasis mine]. You could insist on the genuine propertyhood of E as much as you like, but the victory would be empty. The conclusion, therefore, has to be this: as a significant scientific property, E has been reduced—eliminatively. (Kim, 1999, pp. 18-19).

Kim’s model of reduction is clearly distinct from classical reduction. Additionally, as demonstrated by the example of the gene, it is not excessively unwieldy. It can be readily applied, though the result, as noted by Batterman (2004, p. 70) is “species—or structure—specific”. Thus we have “reductions of human psychology to human physiology (and ultimately to physics) and reductions of reptilian ‘psychology’ to reptilian physiology (and ultimately to physics). We do not, however, have a reduction of *psychology* to physics.” Yet this should not be troubling, unless, like Fodor, one wishes to maintain dogmatically the independence of higher-level scientific disciplines from physics. Attempts at wholesale reduction are, due to the sheer size of the task, almost certainly doomed to fail. Instead, given the complex and interwoven nature of all the material grouped together under broad headings like psychology, piece-wise reduction seems much more likely to succeed.

In addressing why Kim's model is preferred over the classical model, I am not sure that a sufficiently strong answer can be given that would satisfy opponents of reductivism, but all the same I will give my reasons. The primary motivation is the immense difficulty involved in formulating the necessary bridge laws of the classical model. Recall that for reduction to take place under the classical model the higher-level sciences must relate to the lower-level sciences via laws that translate the (scientific) predicates of one into the other. This is a particular problem when the sciences do not share the same terminology, such as—to take a drastic example—the reduction of psychology to physics. The actual path may be indirect, moving from psychology to neuropsychology, neurology, biochemistry, and so on, down to physics, but terminological problems still abound. Each step is laborious to formulate and requires intricate knowledge of not only the laws of each level but also a considerable understanding of how they can be translated. It seems to me that we lack that required understanding, which is also, ironically, on par with the level of knowledge necessary to declare the existence of emergents in that particular domain. Put differently, if we know the details of how higher-level laws can be related to lower-level laws, we should also know whether such a reduction is possible. And, in knowing this, we would already be able to ascertain the existence of any possible emergents.

Kim’s model relaxes these constraints, but not so much as to be useless. The example of the gene, though simplistic, indicates how reduction can begin even in cases where certain theoretical knowledge is lacking. We could have, in principle, functionalized the gene as soon as it was conceptualized as the unit of inheritance, though we would have had to wait a long time for its realizer and accompanying theory to be found. Importantly, however, we do not need to know precisely how the language of genetics can be mapped on to molecular biology; we only need to understand the causal/nomic role played by that which we are functionalizing. Because there is no universal agreement on a concise definition of life (or even whether such a definition is possible or useful), I think it makes sense to employ the model that better accommodates—through functionalization—the spirit of the project, rather than the letter. The correspondence of scientific finding to the functionalized emergent, provided by the second and third steps of Kim’s model, is left to decide the success or failure of the reduction.

The allowance of multiple realizability is another reason to adopt Kim’s model. We will see in the next chapter that life is multiply realized, but the permissibility of multiple realizability in the classical model has been called into question by some, most notably Fodor (1974). Fodor believes that while the laws that hold within the sciences themselves are true laws, the bridge laws used to reduce the special sciences are not. Batterman (2004) gives an example to demonstrate this, which he attributes to Kim. He begins by stating a hypothetical law, that “all jade is green”, and asks whether it can be reduced to the level of its chemical realizers. There are two physical realizers of jade (jadeite and nephrite) and they are distinct physical kinds (Batterman, 2004, p. 65). But “because the microstructures of jadeite and nephrite are completely distinct and heterogeneous from the point of view of the lower level theory, this bridge law cannot express coextensivity of a natural kind term of the jade theory with a *natural kind* term of the lower level chemical theory” (Batterman, 2004, p. 66). Since laws should relate natural kinds to natural kinds, and since a disjunction of natural kinds (e.g. jadeite *or* nephrite) is not also a natural kind, bridge laws cannot be natural kinds.[9] Kim suggests that the way out of this problem is to accept a reduction of jade to distinct instances of jadeite and nephrite, and in so doing dispense with the idea of higher-level laws of science[10]. So while Kim and Fodor are at odds over reductivism—with Kim arguing for, and Fodor arguing against—both agree that multiple realizability presents a significant problem for the classical model’s bridge laws. This problem was, at least in part, responsible for Kim’s development of the functional model, and thus I think it is reasonable to use this model as the means of reducing life to the molecular level.

---

[9] Fodor (1974, p. 102) admits that there may well be irresolvable discrepancies in how people conceptualize natural kinds and laws.

[10] His argument is of course much more detailed and elaborate, but this is his general conclusion.

2.4 Directional Causation

We have now covered the first four tenets of emergentism outlined by Kim. The first two tenets, pertaining to the emergence of complex entities and properties, can be intuitively grasped; explication was required, however, to show that emergent properties should be neither theoretically predictable nor reducible, as necessitated by the third and fourth tenets. This section deals with the final tenet, the causal efficacy of emergent properties, and in particular their required *direction* of causation.

Most accounts of emergentism insist that the properties that emerge have causal powers distinct from the causal powers of their basal conditions. Examination of these novel causal powers leads to a potentially problematic discovery, namely, that there appears to be downward causation. First, though, what does it mean to say that something has downward causation? We have already granted that properties emerge from an emergent base, since emergence is dependent on the complex structural configuration of higher-level entities and their properties (tenets (1) and (2) of emergentism). For that emergent base then there is always an emergent property that is instantiated whenever those basal properties are present—this is simply supervenience. If we want to say that an emergent property has novel causal powers, we must be saying something about the emergent property's ability to cause same-level properties. (The reason we do not need to consider higher-level causation is that the higher level would necessarily have basal properties that must be caused in order for the higher-level property to be caused.) Accordingly, each of these emergent-level properties must have respective basal physical properties upon which they supervene. But because each emergent property's base must be present for the emergent property to be present, it follows that to say that the emergent property caused another emergent property is to say that it caused the second property's basal properties, which in turn caused the emergence of the second property. And this is downward causation, from higher level to basal level. In practice, this amounts to saying that, for example, consciousness exhibits downward causation by altering the world at the physical level (Kim, 2006).

The *problem* that arises should already be evident in the above description of downward causation. If one emergent property causes another, or more precisely if one causes the emergent base of another, and has its own emergent base (as it must), then why can't we simply say that the first emergent base caused the second? Why do we need to talk about emergent properties at all? Kim (2006, p. 558) uses the association of pain and neural states to illustrate: "there is little to recommend in the claim that a neural state causes pain and then pain in turn causes, say, my hand withdrawal. How can there be a causal chain from pain to the hand motion that is separate and independent from the physical causal chain from the neural state to the motion of the hand?" And because we are nomologically required to accept the first set of basal properties as a cause of the second, if we insist on including as an additional cause the emergent property caused by the first set of basal properties the result is causal overdetermination. This puts emergentism in a difficult position: "if there is systematic causal overdetermination in all cases of downward causation, emergents cannot fulfill their causal promise; anything they causally contribute can be, and is, contributed by a physical cause….If downward causation goes, so goes emergentism" (Kim, 2006, 558).

Thus, while downward causation is not itself an incoherent idea, it poses a difficult problem for emergentism. Kim (1999; 2006) insists that there has not yet been an acceptable solution found by emergentists, and this leaves emergence on tenuous footing: on one hand, it must accept (and indeed is predicated on) the notion of downward causation; on the other, downward causation seems to preclude the properties that are proposed to have emerged from having any real causal value. Kim (1999) suggests that there may be value in retaining a *conceptual* interpretation of downward causation that leaves intact the representational features of the higher-level sciences, such as their concepts and descriptions. This would allow for "a single causal relation [to be] describable in different languages"—that is, the languages of the different sciences (Kim, 1999, p. 33).

Stepping back, from considerations found above there are at least two ways in which emergentism may be undermined. First, if any property is reducible to its basal level, it is not emergent but must instead be resultant. I suspect most emergentists will readily grant this, as they are unlikely to insist on the impossibility of reduction *simpliciter*. However, broad claims of the reducibility of all special sciences to physics greatly restricts the domain of possible emergents, leaving in many cases emergents only as epiphenomenal features with little or no explanatory value.

The second way in which emergentism may be undermined is the already-discussed problem of downward causation. Unlike reduction, downward causation is a general and deeply-rooted concern for emergentists. Some properties may be reducible while others are not, and thus there is still room available for emergence; but, if the argument against downward causation holds, it is unclear how *any* account of emergence can proceed without first accounting for how the proposed emergents are able to exert their causal powers.

The goal of the next chapter is not to provide a decisive argument against emergentism as a doctrine—if it were, I would likely employ the problem of downward causation. I want to grant that perhaps there is a solution, and that some clever accounts of emergence may posit a means of causal efficacy that circumvents this particular problem. My focus will instead be on a reduction of life to a basal level, which, while leaving open the door open for other emergent properties, will if successful demonstrate that life is not emergent.

Chapter 3: The Emergence of Life(?)

3.1 Introduction

In this chapter, I will mesh the topics covered in the previous chapters to show that life is not, in fact, emergent. The next section will be concerned with eliminating a possible ambiguity of terminology when talking about emergence and life that may arise as a result of the proliferation of “emergence talk” in scientific literature. It will be important to distinguish this type of emergence from the stricter form employed in this thesis to prevent both confusion and claims of triviality. The third section details my argument against the emergence of life. I focus on two points: the reducibility of life to the molecular level via Kim’s functional model of reduction, and life’s apparent lack of novel causal powers. I will also briefly suggest that in denying the emergence of life I am not also denying the possibility of emergent properties arising as a *result* of life. The final section will handle some objections to my argument, namely the problem posed by forms of emergence that differ from the one outlined in this thesis, and the claim that Kim’s model of reduction can be used to reduce *everything*.

3.2 Emergence Everywhere

> We now see the term [emergence] freely bandied about, especially by some scientists and science writers, with little visible regard for whether its use is underpinned by a consistent, tolerably unified, and shared meaning (and if so what it is). This has created situations in which those discussing emergence, even face to face, more often than not talk past each other. Sometimes one gets the impression that the only thing the participants share is the word "emergence". The intuitive associations this word evokes in us do not add up to a concept robust enough to do any useful work, or even to serve as helpful constraints on a theoretical account of construction of the concept. "Emergence" is very much a term of philosophical trade; it can pretty much mean whatever you want it to mean, the only condition being that you had better be reasonably clear about what you mean, and that your concept turns out to be something interesting and theoretically useful. (Kim, 2006, p. 548).

Kim's remarks indicate the existence of a very real problem when attempting to examine the proposed emergence of a particular entity or property. In this section I want to identify and differentiate a common usage of the term "emergence", found throughout scientific literature on life and its origins, from the kind of emergence that was the topic of the previous chapter. If one reads the relevant scientific literature, a sense of certainty is gained that might lead one to suppose that the emergence of life is old news, something shown to be true long ago and subsequently adopted into scientific canon. But, as Kim astutely observed, emergence used in this context is almost always taken for granted, and almost never scrutinized; it is a stepping stone on the path to other arguments, hardly worthy of closer inspection. This has the effect of leaving readers to intuit their own meaning of emergence, and I suspect that quite often this meaning is simply something like “appeared”—the context rarely requires anything more.

There is, to be sure, a trivial sense of emergence which need not be argued for, when used, for example, in the statement, “At some point in Earth’s history, life emerged.” Here “emergence” often does just mean “appeared”, and since presumably there was a time in Earth’s history when no life existed, life must have, at some later time, formed. This kind of emergence, however, is clearly not the same as the kind discussed in the previous chapter. Recall that to say something is emergent is to say a good deal about that entity or property: it is complex, has (at least one) basal level from which it emerged, and it possesses novel properties with novel causal powers. That is well beyond what is being claimed by saying that between two distinct times something came into existence that did not previously exist.

Trivial emergence does not seem to be what most biologists have in mind when they talk about life. Those who view life as emergent want to assert something *about* life, that is, they want to go beyond the simple fact that life exists. But despite recognizing—in a vague sense, if nothing more—the philosophical overtones of emergence, most authors do not in practice adhere to a robust notion of emergence of the sort outlined in the previous chapter. An excellent example of this comes from Morowitz’s (2002) *The Emergence of Everything*. Morowitz sketches 28 instances of emergence that he feels to be of particular importance in the history of the universe. His selected examples range from “The primordium” to stars, cells, primates, agriculture, and philosophy itself. We can be sympathetic to his project—identifying specific instances of emergence—since producing even a single conclusive example of an emergent would be groundbreaking. Furthermore, his is a reasonable approach to finding emergents: start at the most basic level (the origin of the universe) and, with the benefit of hindsight, search for higher-level entities that seem to resist explanation at the level of their constituents.

Unfortunately, a cursory glance at the cases he includes indicates that something is likely wrong with his conclusions (and, consequently, with his concept of emergence). An introductory chapter acquaints the reader with a brief, somewhat curious history of emergence as philosophical notion. Morowitz mentions most of the requisite features of emergence discussed already in this thesis, such as complexity, irreducibility, and novelty, though causal efficacy is notably absent. However, once he moves on to the instances themselves, it seems as if the details of emergence are lost, replaced by a version of emergence that is neither trivial nor rigorous but instead somewhere in the middle. We see this, for example, in his take on the emergence of planetary accretion. He concludes that particular chapter with the statement, “The emergence of planets in any case is part of the process of star formation, surrounding the stars with potential abodes for life. The rules with which they emerge are pointing to higher degrees of organization” (Morowitz, 2002, p. 62). In his next chapter, on planetary structure, he says,

> There is little to suggest that any principles beyond thermodynamics, physical chemistry, and mechanics are required to understand the structure formation of the Earth shells, but nevertheless some details of Earth formation are still lacking.... Even when the reductionist principles are thoroughly understood, the complexity of the unfolding leaves us with much to understand. This provides another role of emergence as a tool in understanding our world.... The bottom line is that the emergent planet is complex, both structurally and kinetically.... Understanding the structure of the Earth seems to lie within the domain of normal physics and chemistry. It is, however, clearly a problem of great complexity and great difficulty (Morowitz, 2002, pp. 64-66).

In each case, his usage of "emergence" is entirely consistent with roughly meaning only that there is a trend towards increasing complexity. This is certainly a trend he intends to convey, as the examples (chapters) are arranged chronologically to demonstrate that complexity increases with time, building on the (complexity of the) configuration of the level from which it emerges. But he quite clearly violates a basic tenet—irreducibility—of emergentism in stating that the emergent can be understood, though with difficulty, at a basal level. We could charitably interpret this as a slip on Morowitz's part, but I think it it is more profitably conceived as insight into the extent to which he intends the term "emergence" to apply.

Marshaling sufficient evidence for a particular claim of emergence is an exceedingly difficult task, even if one is only concerned with what appear to be straightforward physical events[11]. Morowitz's early topics are of this sort, just as one might expect in a universe that is becoming increasingly complex with time. The progression moves from the inception of the universe through to stellar formation, planetary accretion, geosphere formation, and so on. It is generally accepted that these events have unfolded according to what are now fairly well understood physical laws, making claims of emergence difficult to defend since such claims imply the inability of physics to explain adequately the higher-level features. However, if one is primarily concerned only with showing increasing complexity, combined with acknowledgment of the "novelty" of the entities that arise (without careful consideration of whether the entities are merely resultant), the amount of supporting evidence needed is lessened considerably. And, I think, it is *only* by recourse to this simplified version of emergence that the task of defending the emergence of later topics, such as agriculture and philosophy, are possible at all. His discussion of the emergence of mammals effectively illustrates this point. If we think of emergence only as representing an increase in complexity with an end-product that did not exist previously, the transition from reptile to mammal could reasonably be described as emergence. But to label this evolutionary path truly emergent, it would need to be shown, for example, that the emergence of mammals is neither explainable at, nor reducible to, some basal level (presumably reptiles?). Additionally, mammals would need to possess novel causal powers capable, though its meaning in this context becomes obscure. Instead, an overview of Morowitz's chosen cases of emergence looks more like a series of increasingly complex events with complex explanations that are perhaps most *easily* described using higher-level

---

[11] Batterman's (2004) discussion of emergence with respect to rainbows is instructive on this point.

concepts that do not readily map onto lower-level concepts. What he seems to identify as emergents are often just resultants, and it is only in a strained, weak sense that we can consider something like spirituality or philosophy to be emergent.

There are many sophisticated accounts of emergence in scientific writing, accounts which in ascribing emergence to some property or feature of the world invoke something very much like what I have called a robust notion of emergence. P.W. Anderson's (1972) article *More is Different* is one such piece in which a stronger version of emergence—more robust than Morowitz's emergence, and less than Kim's formulation—is used. Anderson insists that though we may be justified in accepting reductionist hypotheses, we are not therefore entitled to assume that these hypotheses are also "constructionist": "The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe" (Anderson, 1972, p. 393). Anderson uses detailed examples of broken symmetry in many-bodied physical systems to show that at each level in the natural structural hierarchy "entirely new laws, concepts, and generalizations are necessary" (p. 393). Whether one agrees with his conclusions or not it is readily apparent that the kind of emergence he has in mind involves more than just increasing complexity. One particular point of interest in his belief in the reductionist programme, since that seems to defy conventional emergentist thinking. However, close inspection shows that while he accepts reductionism, he still insists that there remains some amount of inscrutability when viewing the world through reduced theories. Theoretical inexplicability (or perhaps inapplicability) of entities and relations using lower-level theories, despite admission of the possibility of reduction, is an interesting and complicated notion. And while it may be somewhat at odds with the received view of emergence, it exemplifies a deeper, more robust form of emergence that calls attention to—rather than glossing over—the problem of finding emergents in physical systems.

Morowitz, Anderson, and others are, of course, free to use “emergence” as they see fit. A reading of *The Emergence of Everything* would likely lead one to believe that emergence is an accepted phenomenon in science and philosophy that occurs whenever a new, complex entity arises. However, if one wishes to convey something of philosophical importance when discussing emergence, one needs to be 1) explicit in its meaning, and 2) consistent in its application. The kind of emergence that Morowitz insists has happened a great many times through the course of time is, thus, not the kind of emergence I have in mind, though Anderson’s emergence is much closer. The moral of this section is two-fold: first, we should not simply declare something to be emergent on the basis of it being novel and more complex than its constituents (or, rather, we can, but at the expense of saying something particularly meaningful); and second, we need to be wary of claims in scientific literature about purported emergents. Though they abound, the strength of the claim of emergence is contingent on the rigor of the author’s definition. Often, as the Morowitz example is intended to show, these definitions are simplified formulations that leave out some key features of a robust definition. In the arguments to follow, emergence will be taken to imply the five tenets outlined in the previous chapter; thus, complexity is certainly necessary, but there is more as well. What will prove particularly important are the points regarding irreducibility and novel causal powers, points that are left out entirely in Morowitz's version of emergence.

3.3 Why Life is not Emergent

Finally, the tools are in place to address the primary question of this thesis: is life emergent? This is a loose, convenient formulation of the issue, but it will serve to move the discussion forward. To begin to answer this question, we need to know two things: 1) we need an operational definition of life 2) we need to know what it means for something to be emergent. The first chapter of this thesis developed the idea that life is best understood as the simultaneous possession by an entity of three features (cellularization, metabolism, and heredity); furthermore, the most basic unit in which these features are found on Earth is the (prokaryotic) cell. The second chapter sorted out the main features of emergentism, based on Kim's five tenets. Of particular importance was the idea of irreducibility, and for that reason two models of reduction were presented. If a proposed emergent resists reduction, it is a good candidate; on the other hand, if the entity or property can be reduced—by appealing to the methods of either model presented, or perhaps others—then it cannot be emergent.

I believe that Kim's model of reduction can be employed to show that life is reducible to its components, and that the joint possession of these components by an entity confers no novel causal powers beyond the causal powers of the basal structure. If this assertion is true, life itself does not seem to be emergent; rather, the predicate *is living* merely expresses a resultant property. Let me show why I think this is the case.

Kim's method requires first that the property in question be functionalized. He used the example of a gene, which was functionalized as, “”the property of having some property (or being a mechanism) that performs a certain causal function, namely that of transmitting phenotypic characteristics from parents to offsprings" (Kim, 1999, 10). Can we do something similar with life? I think we can, based on the definition of life arrived at in the first chapter. It is more complicated than the gene example, because while genes have only (roughly) one causal function, there are three for life. The functionalization of life looks something like this (using the same structure as for the gene):

> Life is the property of having the joint properties that perform certain causal functions, namely that of undergoing metabolic reactions, replicating with both heredity and variation, and isolating the system from the surrounding environment by a selectively permeable barrier.

Is this adequate as a functionalization of life? Kim says that "E's being instantiated is for a certain property P to be instantiated, with this instantiation bearing causal/nomic relations to the instantiations of a specified set of properties in the base domain" (Kim, 1999, p. 11). By substituting into Kim's generalized comment life in place of E and the three properties comprising life in place of P, we seem to be on the right track. For example, cellularization (functionalized as the property of isolating the system from the surrounding environment by a selectively permeable barrier) can be grasped at the level of the components of this barrier. If we wish to label this base domain, it can be identified with molecular biology. The constituents of the barrier are, recall, fatty acids, sometimes interspersed with membrane proteins, that self-assemble into a spherical formation, effectively closing off the contents of the sphere from the outside. At the molecular level—that is, at the level of the constituents of the barrier—certain molecules are able to enter the cell, while others are held outside. This serves many purposes, one of which is to set up ion concentration gradients, which then further control many aspects of cellular homeostasis. Internal cellular (osmotic) pressure can be regulated, hormones can selectively enter the cell to trigger or alter gene expression, and so on. At the molecular level, the cellular membrane is a hot-spot of activity. And the same can be said of the remaining two features of life, each being understood independently at the level of molecular biology.

The second step is to identify realizers of life. Once again, remember that for life to be instantiated, cellularization, heredity, and metabolism must be instantiated. At the macroscopic level this is a fairly easy task. Simple observation led to the conclusion that humans and animals eat and breathe, procreate, and are differentiated from their environment. However, to understand these processes at the level of (what would become) molecular biology took many years, and in some respects it is still an ongoing project. Furthermore, cells, for example, do not carry out these tasks in the same way as macroscopic entities. And insofar as we are concerned with life emerging from non-living molecules, it is the microbiological level of understanding that is necessary. As we might expect, it turns out that there are an immense number of possible realizers of the features of life, though this depends to some extent on how one identifies these realizers. In the case of the gene, DNA was found to be the realizer, but we do not count each DNA molecule as independent in this respect; DNA identifies a type of realizer for the gene (the present one), RNA another, and so on[12]. Likewise, it seems wrong to count each individual as a realizer of the properties of life, since many individuals share the same biological processes. It is even unclear whether the line should be drawn at different species; for example, there are a universal set of metabolic pathways, passed down through evolution, with each species manifesting some particular subset that best suits their needs. Many species share the same subset of metabolic pathways. Thus, to say that there are 10 million extant species is not to say that there are 10 million different realizers of metabolism. And given that the vast majority of extant species are unicellular (prokaryotic), the shared means by which they isolate themselves from their environment guarantees a considerably lower number of realizers of cellularization. But we need not be concerned with the exact number of realizers for each feature; there are *some* realizers, and each can be understood individually. Patterns that arise simply make the process of scientific investigation less tedious.

The final step is to find a theory, at the basal level, that explains how the realizers work. As predicted this goes hand-in-hand with identifying the realizers themselves. It is not enough to know only that fatty acid chains make up cellular membranes: we also need to know how they do it. With respect to the features of life, we do indeed have the required theories at the molecular-biological level. For cellular membranes, we understand not only how phospholipids self-assemble to form spherical vesicles, but we also understand how concentrations gradients and ATP-powered pumps are used to

---

[12] Particular genes are realized by certain configurations of DNA, though in the past they may also have been realized by different configurations, or by a different ribonucleotide altogether, such as a certain sequence of RNA. Here I am not concerned with *particular* genes but instead the more broad means of inheritance characterized by, say, the entire human genome.

allow selective entry to cellular nutrients. There are too many specific functions of cellularization to list exhaustively, and perhaps some remain to be discovered, but based on our understanding of its current functions it seems reasonable to think that whatever new functions are discovered will also be comprehended equally well at the molecular level. And we can say the same for the remaining two functions, because we know how DNA is able to code and pass along genetic information, and how the reactions in the metabolic pathway proceed.

With functionalizations of the three features of life in hand, along with considerable scientific evidence for there actually being realizers (and theories about how those realizers work), we are entitled, according to Kim's model, to say that life has been reduced. And, as irreducibility is a key tenet of emergentism, a property that can be reduced cannot be considered emergent. Therefore, from the point of view of a robust form of emergence, requiring (amongst other things) that emergents be irreducible, life—characterized by the three necessary features of cellularization, metabolism, and heredity—should not be seen as emergent.

I also expressed my belief that life *qua* biosystem does not exhibit any novel causal powers. I think this is apparent if the reducibility of life is conceded, but it is also an independent empirical question. If life is reducible, as I have suggested above, then the causal powers of life (and in particular, elementary biosystems, or cells) can be understood at the molecular level too. This follows for what we might call "resultant causal powers", as opposed to novel causal powers. A resultant causal power could be something like the paperweight's ability, mentioned in the previous chapter, to hold paper to a desk. Another example, mentioned by Kim (1999) is that of a bird, whose individual parts (simplified to consist only of head, wings, and so on) are capable of a kind of movement that the individual parts cannot undergo alone. So, if life is reducible to the molecular level, and if in doing the reduction we are required to understand the features of life at the basal level, then the higher-level features we observe (and which we can reduce) are merely resultant. I think this point will be accepted by most emergentists, who will choose instead to take issue with the reduction itself.

But the empirical question still looms: are there in fact emergent properties? Here there seems to be a good deal of room for dispute, for emergents cannot (by definition) be explained at, or predicted by knowledge of, their basal level. In scientific investigation, one frequently comes up against the limits of current technology or theoretical explanation, yet higher level entities and properties clearly exist, and we endeavor to understand how they come about. This creates explanatory gaps, gaps between the lower levels and the higher levels, and between sub-levels within levels. Much hangs on how one interprets these gaps. Emergentists insist that some of these gaps are brute fact, and that even a complete understanding of the lower-level science will not be sufficient to explain the complex properties and/or entities that emerge. The problem is, *at what point should science cease investigation and admit emergence as a brute fact, rather than holding out for better technology and theoretical or conceptual understanding?* In the case of life, we seem to understand its processes at the molecular level. As an elementary biosystem, we can describe the cell at the cellular level and examine its internal and external interactions, or we can look at the molecular picture. But is there any evidence of causal powers that exist at the level of the cell that *cannot* also be explained by our knowledge of its molecular workings? Science, to my knowledge, has found no such thing. There are to be sure *resultant* properties and entities, but so far nothing more.

If we widen the gap by looking not between the levels represented by cellular and molecular biology but instead by setting as the reduction base (bio)physics, the reduction becomes immensely more complex. We are required to span a critical divide *between* levels, crossing from the realm of biology into physics. The way forward is not entirely dark: some inroads have been made into comprehending biological processes at the level of physics. But we are, essentially, left hanging at the steps in Kim's reductive model that require knowledge of the realizers and how they work, because we cannot offer a full theory at the physical level to account for complex biological interactions.

Now, this is precisely what emergentists want to deny. They see this not as an inadequacy on the part of science but as a fundamental limit to the explanations of a given level (physics in this case). We can and should study the base (physics) and the higher-level theory (biology), but we must accept that the former cannot fully explain the latter. Yet, to the scientific mind, I think there is something deeply unsettling about the suggestion that there are limits to enquiry and explanation that end in brute facts. Developments in biophysics have clearly demonstrated that attempts to understand biology through physics are not destined to fail. There appears to be a great deal left to understand, and scientists, knowing this, press on. We are, I think, far from possessing the level of knowledge necessary to provide full theoretical descriptions at the base (physical) level of all things biological, but that same lack of knowledge should caution us against declaring too soon the existence of fundamentally inexplicable higher-level facts.

This seems to be the greater problem when one considers reduction of biology; critics of reduction are less concerned with intra-biological reduction, and more concerned with the (im)possibility of reducing biology to physics. And while I want to acknowledge that it is indeed a tremendous theoretical hurdle, it is too early for emergentists to claim victory. Fortunately, by limiting the reduction base to molecular biology we are able to say with much greater confidence that there is no remaining way in which novel causal powers of cells may be manifested. As a result, we are (inversely) proportionately justified in asserting molecular biology’s ability to fully explain elementary life. Ideally, once we have the theoretical knowledge required to design comprehensive (and exhaustive) searches, the existence of emergents (or lack thereof) will be settled decisively by empirical evidence. In the meantime, however, it seems unwise to ascribe with any surety emergence to entities and properties that arise from somewhere in the gaps of our knowledge.

Before moving on to objections, I want to insist that I am opposed here only to life itself (as elementary biosystems) being considered emergent. Specifically, I do not wish to rule out the possibility of emergent properties arising at some point as a result of life.

If there are emergent properties at all, first-person conscious experience seems like a good candidate (Kim, 1999). Thus, it may be the case that where life develops into sufficiently complex organisms, emergent properties arise. Importantly, however, this fact, if true, does not require life to be emergent as well. For emergentists who wish to maintain life's emergent nature, they must insist on life emerging from abiotic material. Emergence from a collection of non-emergent basal properties is not, then, a foreign idea. So, it seems reasonable to assume that whether consciousness is emergent depends in no way on life being emergent. The existence of life certainly represents a high degree of organizational complexity at the molecular level—this, perhaps, is why some people mistakenly assume life to be emergent, by equating emergence with complexity, as discussed in the preceding section. But what rises out of this structural complexity does not—and maybe *cannot,* on a strict reading of emergentism's tenets—depend entirely on the (non)emergence of the basal properties and the laws that govern them.

3.4 Problems

I do not doubt that any nonreductive physicalist worth her salt would object in some way to the preceding line of argument. Batterman (2004, p. 70) suggests that this may simply amount to a digging in of one's heels in resistance to any form of reduction based on the desire to "maintain orthodoxy expressed by the label 'nonreductive physicalism'." However, there are some legitimate concerns that have not yet been addressed, and in this section I will attempt to remedy that problem.

One obvious problem is already suggested in Anderson's take on reduction: what about emergentists who deny that reduction is a problem while still insisting on the existence of emergents? And, for that matter, what about entirely different characterizations of emergence? I intended to head off, to some extent, this line of objection in the first section where I considered simplistic forms of emergence, especially those that essentially equate emergence with increasing complexity. As noted there, there is no true consensus on what constitutes emergentism. Kim's summary of the received view is a sort of survey of various forms of emergentism that attempts to capture their common features; the five tenets he identifies are not therefore necessarily common to *all* forms of emergentism. The argument provided above for the reduction of life to molecular biology according to Kim's functional model of reduction would certainly run into problems if one's view of emergence did not require irreducibility. This is what we find in Anderson's discussion of emergence, where he accepts reductionism but denies that the reductionist programme offers anything constructive. I think there is some room here for simply agreeing to disagree about what constitutes an adequate characterization of emergentism, but I think more can be said in this particular case.

Anderson's diagnosis of reductionism failing to offer anything constructive seems intuitively correct, but I'm not sure that it does as much work as he hopes. If we take an extreme example, it is patently absurd to attempt to explain something like the subduction of tectonic plates by appealing to particle physics. Yet we can be a reductionist about geophysics, maintaining that tectonic plates can, if we so desired, ultimately be reduced down to their atomic underpinnings. Anderson would, I think, agree with this too, though he would deny that one could then build the requisite theoretical framework from particles back up to plates—this is what his examples of broken symmetry are intended to show. The salient point is that whether or not we can do this, *is it ever desirable to do so?* It seems much more likely that, reducible though tectonic plates may be, it is considerably more convenient to maintain the concepts of geophysics and theories of geophysics with the implicit understanding that those same concepts and theories are actually reducible. Why, then, should we expect any reductionist programme to be accompanied by a constructionist one? Reductionism in a moderate sense is not intended to be eliminative, and conservative reduction (in which the concepts and theories are kept) seems perfectly capable of making the necessary allowances for the convenience of hierarchical terminology while acknowledging the primacy of physics. The need for a constructionist approach is removed if we are always allowed to work backwards, starting with higher-level understanding (because it is, for example, often computationally more manageable) and progressing towards increasingly complex understanding at the basal level, whatever it may be. We do not, however, need to concede (as Anderson insists) that there is an unbridgeable divide in the reconstruction, were we inclined to attempt it. It is undoubtedly unwieldy, and almost certainly beyond our technological and theoretical capabilities at present, but should not on this basis conclude that it is *impossible*.

More generally, some forms of emergentism may dispute other tenets. For example, Batterman (2004) insists that emergents need not necessarily have novel causal powers, because there are some situations in which it makes little sense to talk in this way. Instead, he insists that emergents should factor into novel *explanations*. But defending against every unique claim put forth by emergentists would yield at best only a pyrrhic victory, consuming too much time for too little theoretical gain. A line must be drawn somewhere, and the tenets identified by Kim seem to be a reasonable characterization of emergentism that remains true to what founders such as Broad and Alexander had in mind.

There are other problems related to this approach, such as the general claim that Kim's model virtually guarantees reduction, and that for this reason it is unlikely to be accepted by many nonreductive-physicalists (Batterman, 2004). We can see Batterman's point best in the first step of Kim's model. In functionalizing E, it is automatically related to the reduction base by virtue of the properties in the base. The concern, I think, is that even ardent emergentists will allow for supervenience—that is, a deterministic relationship between certain properties in the reduction base and E. Wherever these basal properties occur, we expect to find E. But emergentists want to maintain that something *more* occurs too. Certainly E supervenes on properties or conditions in the emergent base, but E is also a “brute fact” that cannot be explained or understood completely by knowledge of the emergent base. Yet Kim’s model allows for such reductive leaps by identifying the causal powers of the physical base with the causal powers of the properties that arise. Somehow, mysteriously, emergentists want to break this causal chain (or perhaps just weaken it) to include emergents with their own novel causal powers. But Kim insists that where there are emergents there are physical realizers, and where there are physical realizers as causes we need only speak of the non-mysterious causation between physical realizers. The upshot is that, if there are emergents, they are considerably more difficult to identify than was once thought; the haphazard labeling of many complex systems as emergent was mistaken. Kim’s model puts the burden of proof back on the emergentists by showing many properties and theories to be reducible to the level of their constituents, effectively raising the bar for emergence.

Conclusion

The basic ideas of emergence can be intuitively understood, yet they also seem to go against the grain of scientific investigation. It is almost certain that we will some day understand much of the macroscopic world at the microscopic level; the question is, will we understand it *completely*? Or will some aspects, being emergent, be forever beyond explanation at the level of physics (and instead simply be brute, contingent facts)? And of equal interest is the question whether, in cases where reductive explanation is possible, should the higher-level terminology be dispensed with altogether, or will it still serve some useful purpose? Despite our knowledge of life at both the cellular and molecular level it is often still convenient to use the higher-level descriptions, even when these descriptions convey nothing new; maybe this will always be true, that there is some positive value in the higher-level concepts regardless of their theoretical reducibility. But putting too much stock in these concepts, by, for example, asserting their existence as something that transcends their constituents, is something that, until empirically decided otherwise, should be avoided. In the case of life, there is as yet no reason beyond its (generally) increasing complexity to think it is emergent. Its features are demonstrably reducible to their molecular underpinnings; furthermore, when these features are collected together to form life, what follows is merely resultant, not novel. The claim of novelty, and indeed the claim of emergence, is a positive assertion, but there is nothing to point to in support of these claims; they appear instead to be based primarily on gaps in our current evidential base. And so it is for these reasons that I take life not to be emergent.

References

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

Batterman, Robert. 2004. *The Devil in the Details*. Toronto, ON: Oxford University Press.

Broad, C. D. 1925. *The mind and its place in nature*. London: Routledge & Kegan Paul.

Eigen, Manfred. 1992. *Steps towards Life*. New York, NY: Oxford University Press.

Fodor, Jerry. 1974. Special sciences, or the disunity of science as a working hypothesis. *Synthese*, 28: 97-115.

Fodor, Jerry. 1997. Special sciences: Still autonomous after all these years. *Philosophical Perspectives,* 11: 149-163.

Hurst, C. T. 1937. On the origin of life. *Bios*, 8(4): 201-209.

Karp, Gerald. 2005. *Cell and Molecular Biology: Concepts and Experiments 4th ed.* Hoboken, NJ: John Wiley & Sons.

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

Kim, Jaegwon. 1999. Making Sense of Emergence. *Philosophical Studies,* 95: 3-36.

Kim, Jaegwon. 2006. Emergence: Core ideas and issues. *Synthese,* 151: 547-559.

Mahner, Martin, and Bunge, Mario. 1997. *Foundations of Biophilosophy.* New York, NY: Springer.

Maynard Smith, John, and Szathmáry. Eörs. 1999. *The Origins of Life: From the Birth of Life to the Origins of Language.* Toronto, ON: Oxford University Press.

Morowitz, Harold. 1992. *The Beginnings of Cellular Life.* New York, NY: Yale University Press.

Morowitz, Harold. 2002. *The Emergence of Everything*. Toronto, ON: Oxford University Press.

Popa, Radu. 2004. *Between Necessity and Probability: Searching for the Definition and Origin of Life*. New York, NY: Springer.

Schejter, Abel, and Agassi, Joseph. 1994. On the definition of life. *Journal for General Philosophy of Science*, 25: 97-106.