

1

The Enduring Legacy of Aristotle: The Battle over Life as Self-Organization or (Genetic-Based) Reproduction

1.1 Overview

There are universal theories in physics and chemistry but no universal theories in biology. The failure of biologists to come up with such a theory is not due to a lack of effort. Philosophers and scientists have struggled to formulate universal principles of life since at least the time of Newton. This chapter traces the history of these efforts back to their roots in the work of the ancient Greek philosopher Aristotle. Aristotle's influence can be seen today in the view, which dominates contemporary biological thought about the nature and origin(s) of life, that the following abstract functional characteristics are basic to life: (1) the capacity to self-organize and maintain self-organization for an extended period of time against both external and internal perturbations and (2) the capacity to reproduce and (in light of Darwin's theory of evolution) transmit to progeny adaptive characteristics. For the sake of simplicity, I refer to the former as "O" and to the latter as "R" throughout this chapter. As Section 1.2 discusses, the conceptual parallels between O and R and Aristotle's ideas about life are remarkably close. He identified "nutrition" and "reproduction" as the basic functions of life and debated (as do so many contemporary researchers) which is more basic. Aristotle also bequeathed to biology the thorny problem of teleology – the notion that the allegedly basic functions of life (in their contemporary guise, metabolism and genetic-based reproduction) require a strange (to the modern scientific mind) form of causation that is intrinsically directed at achieving a future goal. As Aristotle argued, living things are not just fed, they feed *themselves*, and they are not just copied, they reproduce *themselves*. Characteristic O reflects this view in explicitly referring to the idea of self-organization. Similarly, characteristic R implicitly assumes that organisms contain an internal principle for generating organisms resembling themselves; external processes do not (like a 3D printer) duplicate them.

Sections 1.3 and 1.4 trace how Aristotle's seminal ideas about life evolved into characteristics O and R. As will become clear, the idea that the causal processes responsible for the distinctive functions of life are intrinsically goal directed poses special challenges to scientific reasoning. Teleological causation is not easily accommodated within the framework of classical physics, which holds that causes precede their effects without anticipating them. With the advent of Darwin's theory of evolution by natural selection, many biologists became convinced that the *prima facie* teleological properties of life could be explained in terms of undirected cause and effect relations after all. As we shall see, this claim is too sweeping. But even supposing that teleological causation has been exorcised from biology, the conceptual parallels between Aristotle's ideas about life and characteristics O and R remain uncannily close. From a philosophical perspective, the parallels between Aristotle's ideas and characteristics O and R are intriguing because, as the history of science reveals, in other domains of scientific inquiry (e.g., chemistry and physics) the most rapid advances occurred after the abandonment of Aristotelian concepts and principles.

The second half of the chapter (Section 1.5) explores the Aristotelian character of contemporary scientific thought about the nature and origin(s) of life. Theories of life are commonly articulated in the form of definitions. Following in the footsteps of Aristotle, definitions of life invariably privilege one of the characteristics O or R over the other (Section 1.5.1). Those privileging characteristic O (metabolic definitions) struggle with explicating a philosophically coherent and scientifically fruitful concept of self-organization, frequently falling back upon concepts (e.g., autopoiesis) that sound suspiciously teleological.

Theories (often called "models") of the origins of life divide along the same lines as definitions of life with regard to characteristics O and R (Section 1.5.2). Tacit appeals to causal processes that are self-generating and goal directed are routine, for example, the "spontaneous assembly" of chemically improbable, primordial biomolecules, such as small proteins (peptides) or small RNA molecules, from more basic molecular components, and the "emergence" of proto-organisms from complex, autocatalytic, chemical reaction networks. In other words, not only do O and R closely parallel Aristotle's account of life in terms of nutrition and reproduction, they are also difficult to explain in terms of ordinary, undirected causal processes. Some origins of life researchers attempt to circumvent this difficulty by arguing that such events are merely extraordinarily improbable, as opposed to being the products of a weird form of causation. As we shall see, however, this strategy is just as unscientific as one that embraces teleological causation.

The purpose of this chapter is to motivate the thesis that contemporary thought about life may be being held hostage to neo-Aristotelian ideas.¹ For this reason it covers a lot of ground (historical, philosophical, and scientific), much of which will be revisited in greater depth in subsequent chapters. Let me hasten to add that I am not claiming that a neo-Aristotelian framework for theorizing about life is mistaken. Although it has thus far been unsuccessful in providing a general theory of familiar Earth life, let alone life considered generally, and is fraught with conceptual problems, it might still turn out to be the best approach. I am arguing that we should stop *reflexively* viewing life through Aristotelian lenses. Before giving up on the prospects for universal biology, as so many pluralists recommend (see Chapter 6), we need to explore the possibility that the ostensible lack of unity among biological phenomena is the result of tacit commitment to a defective, neo-Aristotelian, theoretical framework for reasoning about life. For as Chapter 4 discusses, an unpropitious set of basic theoretical concepts can frustrate the search for unity among a diverse body of natural phenomena *even when it exists*; not every way of carving up a domain of natural phenomena into fundamental categories is capable of yielding empirically powerful, unifying generalizations.

1.2 Aristotle on the Nature of Life: Nutrition Versus Reproduction

The history of thought about the nature of life is rich and complex. I cannot do it justice here. My purpose is to trace the development of a few, highly influential, core assumptions about life in a selective manner.² Chief among them is the idea that the distinctive features of life are the product of a special form of causation not found in inanimate material systems. This supposition is important because, as we shall see, it underlies functional characteristics O and R and, most importantly, has reappeared in contemporary scientific debates over the nature and (especially) the origins of life. Our historical journey begins in ancient Greece with the writings of the philosopher-scientist Aristotle. As will become apparent, in almost every domain of natural science except speculation about the nature and origins of life, Aristotle's views have been rejected.

Aristotle (384–323 BC) is credited with being the first biologist because he emphasized the importance of basing theoretical conjectures about life upon empirical investigations.³ Although limited to what could be seen with unaided human vision (primarily large, complex organisms, viz., plants and animals),

¹ A recent resurgence of neo-Aristotelian ideas about life among philosophers (e.g., Bedau 2010; Groff and Greco 2013) illustrates the continued influence of Aristotle's ideas.

² See Coleman (1977), Grene and Depew (2004), and Sapp (2003) for detailed and authoritative discussions of the historical development of biology.

³ See Guthrie (1990, Ch. II) for more information on the life of Aristotle.

Aristotle engaged in detailed and systematic investigations of a wide variety of organisms. He observed animals in their natural habitats, studied their development, experimented on them, dissected them and even vivisected them. For these reasons, he is widely and appropriately known as the father of biology. During a period of self-imposed exile on the island of Lesbos, Aristotle spent several years conducting careful studies of marine animals in tide pools. Here he met and collaborated with Theophrastus, a native of the island, who was interested in plants. Aristotle invited Theophrastus to join him in Athens at the Lyceum, where they continued their collaboration, with Theophrastus focusing on plants and Aristotle on animals. Many of Aristotle's works on biology (*History of Animals*, *Parts of Animals*, *Generation of Animals*, as well as shorter essays on animal locomotion, respiration, etc., and the highly theoretical *De Anima*) survive to this day, providing a rich source of information about his views about life; see Barnes's *The Complete Works of Aristotle* (1984).

In his most theoretical work in biology, *De Anima* (trans. Barnes 1984 and also Shields 2013; reprinted in Bedau and Cleland 2010), Aristotle discusses the nature of life, distinguishing living things from nonliving things in terms of functional capacities such as nutrition, reproduction, sensation or perception, locomotion, and thinking (*De Anima* Bks I and II). He identifies nutrition and reproduction as more basic than the others. For Aristotle, nutrition is not nutrients (food) but rather the *internal* capacity or "power" [*dynamis*, in ancient Greek] of an organism to acquire (absorb), process (digest), and use nutrients for biological ends such as development, growth, maintenance, and repair; tellingly, he uses the terms "self-nutrition" and "nutrition" interchangeably in his writings. In *The Parts of Animals* (Bk I 639^{b15}–639^{b32}), Aristotle draws an analogy between a builder constructing a house and the development of an organism, arguing that the causal processes involved can only be understood in terms of their end products, respectively, a finished house and a mature organism; the activities of both the builder and the organism are prearranged in such a way as to bring these ends about. He also observes that there is a significant difference between "productions of nature" and the constructions of human beings. Buildings do not construct themselves. The causation involved is external to the material used to construct them, residing in the intentional (according to a plan) causal activity of human beings. In contrast, the causal processes responsible for constructing and maintaining an organism are internal; organisms build and regenerate *themselves*. It is as if the end product (a mature, thriving organism) is already present in ghostly form, actively organizing and guiding the complex causal processes required to produce and maintain it. Such processes are difficult to make sense of in terms of the mechanistic conception of causation familiar from classical physics. Causes of the latter sort (e.g., wind) do not anticipate or somehow

represent (analogous to the plan of a human builder) their effects (falling leaves) in advance of their occurrence.

To account for the distinctive functional characteristics of life, Aristotle postulated a special form of natural causation (*De Anima* Bk I). Living things are *self-causing* in the sense that they have *internal* regulative principles or “souls” (*De Anima* Bk I.4).⁴ The acquisition of a soul by an inanimate object creates an “organized being” having the capacity to “move” its parts (“organs”) in such a way as to engage in a certain type of functional activity (nutrition, reproduction, locomotion, etc.) as a whole (*De Anima* Bk II 412^{a28}). Nonliving phenomena are not self-causing. Lacking the restorative nutritive power, a mountain cannot sustain its massive form against the processes of erosion; subject to the action of wind and water, it becomes smaller and smaller, eventually disappearing. Analogously, lacking the nutritive power, the maintenance of buildings requires the highly organized, external activity of human beings. Living things are different. They have causally efficacious, internal principles of organization (“soul powers”).

Whereas modern science accepts only one form of causation, Aristotle distinguished four: efficient, material, formal, and final (*Physics* Bk II 1–8). The only one with an analog in modern science (and only imperfect or approximate at that) is efficient causation, which proceeds linearly in time from cause to effect. Final causes, in contrast, act in the opposite direction, shaping and directing occurrent processes towards future goals that are somehow anticipated in advance.⁵ To the extent that the activities generated by a soul power are directed at achieving a certain end, it is both an efficient and a final cause. As an illustration, consider the development of a fertilized egg into a mature organism. That a particular zygote will develop into a zebra – as opposed to, say, a frog – is predetermined. The causal processes involved unfold as one stage (efficiently) causes the next, but always in such a way as to (finally) produce an adult zebra. Importantly, there is no *external* guiding hand. It is as if the zygote contains within itself a self-following recipe for making a zebra.⁶

The most important of the life functions (nutrition and reproduction) are unambiguously teleological. The internal activities that realize them are directed at

⁴ Aristotle’s concept of “soul” (or *psuchê*, in ancient Greek) is very different from the modern concept. A soul is not a divine entity that survives death, nor does it presuppose consciousness or intelligence. It is a principle (or, as it is translated, “breath”) of life, i.e., a cause or power that confers a life function on a material thing (Claus 1981). According to Aristotle, different life functions are associated with different souls.

⁵ Some scholars of Aristotle (e.g., Wieland 1975) contend that his “four causes” are not genuine causes but rather something closely related, viz., forms of explanation. For our purposes, however, we do not need to settle this issue about Aristotle. The important point is that neo-Aristotelian accounts of how life differs from nonlife interpret them as causes, distinguishing final (goal-directed) causation from efficient causation. Much of the modern literature on emergence, for instance, is dedicated to fleshing out a coherent concept of how a cause can be goal directed.

⁶ Viewed in light of what we have learned about how hereditary information is stored on DNA and translated into proteins for use in constructing organismal bodies, Aristotle’s account sounds remarkably contemporary!

bringing about particular ends, for example, a mature, thriving organism. But, while self-generated activities and goal-directed activities often coincide, they need not. Indeed, for Aristotle, these functions may come apart in the case of locomotion (Furley 1994).⁷ The locomotion of an antelope, for instance, could be just for the sake of moving; it need not have a definite purpose, such as fleeing from a lion or seeking food. Nevertheless, the motion of the antelope is self-caused in the sense that its causal source lies *within* the antelope, arising from the internal coordination (or, in Aristotelian terminology, “movement”) of nerves, muscles, and sinews. Thus, just as being goal directed is no guarantee of being self-caused (e.g., the construction of a building is goal directed but not self-caused) so it seems that being self-caused is not, for Aristotle, a guarantee of being goal directed. Aristotle’s account aside, however, there is a weaker sense in which the internal activities of a running antelope are surely goal directed: Their “purpose” is to generate *spontaneous* motion in the antelope. In contrast, the activities that inanimate objects participate in are neither goal directed nor self-generated. A stone is incapable of spontaneous motion; only an external cause, for example, rapidly flowing water or the motion of a foot, can move it. We will return to the distinction between being self-caused and being goal directed shortly. As will become apparent, some of the contemporary debates over the nature and origins of life implicitly appeal to it.

Impressed by the hierarchical functional organization of the parts of animals and plants, Aristotle classified organisms in terms of combinations of life functions. All living things (plants and animals) share the most basic soul powers, (self-) nutrition and reproduction. In this context, it is important to distinguish Aristotle’s concept of reproduction from his notion of *regeneration*, which he would classify under nutrition because it involves the capacity of an existing organism to sustain and repair itself. Reproduction, in contrast, is concerned with the internally orchestrated “production of another like itself” for the purpose of preserving the species of which it is a part (*De Anima* Bk 2.4 415^{a2}; trans. Barnes 1984, 661). Some living things have additional life functions and can be classified taxonomically according to which ones they have and which ones they lack. As an illustration, animals have sensation and humans have, in addition to the soul powers of animals, the capacity for thought (*De Anima* Bk 2.3 414^{a30}–414^{b19}; trans. Barnes 1984, 659). On Aristotle’s account, the other life functions presuppose reproduction and nutrition, for example, sensation cannot occur without nutrition.

⁷ For Aristotle, self-causation amounts to internally generated “movement” of the parts of an organism, which gives rise to vital functions such as nutrition and locomotion (*De Anima* Bk 1.4). It is thus important to distinguish Aristotle’s concept of locomotion (change in place of the whole organism) from his concept of movement (*qua* self-causation).

Scholars disagree about whether Aristotle believed that one of the basic life functions is more fundamental than the other. In some passages he suggests that nutrition is more fundamental: “It follows that first of all we must treat of nutrition and reproduction, for the nutritive soul is found along with all the others and is the most primitive and widely distributed power of soul, being indeed that one in virtue of which all are said to have life” (*De Anima* Bk 2.4 415^{a22}–415^{a26}; Barnes 1984, 661). Yet, as Gareth Matthews (1992) observes, there are also passages in which Aristotle suggests that reproduction is the more fundamental life function because “the production of another like itself . . . is the goal towards which all things strive, that for the sake of which they do whatsoever their nature renders possible” (*De Anima* Bk 2.4 415^{a28}–415^{a32}; Barnes 1984, 661). Matthews concludes that Aristotle is best interpreted as holding that “generation” (reproduction) is more essential to life than nutrition, and moreover that he literally “defines” life in terms of generation. Aristotle’s writings support this view: He speaks of “defining” the different types of soul (which, as mentioned above, distinguish living things from nonliving things) in a manner analogous to the geometer’s paradigmatic definitions of ‘square’ and ‘triangle’ (*De Anima* Bk 2.3 414^{b20}–414^{b33}; trans. Barnes 1984, 659). Nonetheless some scholars of Aristotle, for example, Shields (2002), contend that he was not concerned with defining life.⁸

For our purposes, however, the important point is that there are close conceptual parallels between Aristotle’s functional characterization of life, in terms of self-nutrition and self-reproduction, and contemporary characterizations of life, in terms of O and R. Aristotle explicated self-nutrition metabolically, in terms of the internal capacity of organisms to absorb and process “food” found in the environment for the purpose of bringing about and “. . . maintain[ing] the [organized] being of that which is fed” (*De Anima* Bk 2.4 416^{b1}–416^{b17}; trans. Barnes 1984, 663). He characterized reproduction as species preserving to the extent that it involves an organism generating another organism that closely resembles it. Indeed, as Ernst Mayr (1988, 55–60) observes, Aristotle’s discussions of reproduction, development, and adaptation in living organisms sometimes sound uncannily modern. What does not sound modern is the idea of goal-directed self-causation as the ultimate explanation for the difference between living and inanimate material things. Yet as adumbrated above, and discussed in greater detail below, it is difficult to articulate characteristics O and R without at least tacitly invoking it. The question is can this unsettling form of causation be expunged from O and R?

⁸ I find this view implausible. Aristotle founded the discipline of logic, and wrote extensively about the logical character of definition, which suggests that he would not use the Greek word for definition sloppily. Still, I am not a scholar of Aristotle, and there is some disagreement over this issue. (Perhaps, and this is pure speculation, “nutrition” captures self-causation and “reproduction” the goal-directed aspects of life; since, as discussed earlier, the two concepts are not the same.)

1.3 Classical Mechanism about Life: From Optimism to Quiet Desperation

Aristotelian ideas explicitly dominated scholarly thought about life until the writings of René Descartes (1596–1650).⁹ Influenced by the rapidly developing mechanistic account of nature, which culminated in the seventeenth century with Isaac Newton’s laws of motion. Descartes argued, in “Treatise of Man” (trans. Hall 2003; reprinted in Bedau and Cleland 2010), that living things are automata, albeit much more complex than the artificial devices powered by water, air, and wound springs popular during his day; the bones, muscles, and organs of animals were likened to cogs, pistons, and cams. The only analog to Aristotle’s four causes in Newtonian physics is force (efficient causation); material causation is relegated to passive matter, and formal and final causation are completely absent.

But in the ensuing years, as scientists learned more about plants and animals, it became clear that organismal bodies differ in important ways from mechanical devices. Vitalism, the view that life involves a special kind of force, energy, or substance, was a popular response to these difficulties.¹⁰ The most influential versions of vitalism (e.g., that of Jöns Jakob Berzelius) attempted to accommodate life within Newton’s highly successful theoretical framework by postulating a special life force for maintaining the dynamical functional organization of living things. Such an approach seemed eminently reasonable. Newton’s three laws of motion applied to forces considered generally. They left open the identity of particular forces, including the possibility that new forces would be discovered.

Worries about whether life could be accommodated within the theoretical framework of classical physics culminated in the late eighteenth century with the writings of Immanuel Kant (1724–1804), who famously concluded that there would never be a Newton of biology (*Dialectic of Teleological Judgment* §75; trans. Guyer and Matthews 2001; reprinted in Bedau and Cleland 2010). Kant argued that life could only be understood on the assumption that it is produced according to a design; organisms are intrinsically purposive or goal-directed entities. The source of this design, however, is not an intelligent designer but rather nature.¹¹ How is it possible for design to exist in nature without an intelligent designer? According to Kant, the very concept of something being a natural end requires conceiving of it as self-causing, that is, as “both cause and effect of itself” (§64, 371). Kant also argued that life is unique in being naturally teleological. In short, for Kant, organisms differ from inanimate objects in having internal, purposive, regulative principles that cannot be explained within a

⁹ See Grene and Depew (2004) for more on the evolution of thought about the teleological properties of life from Aristotle to contemporary times.

¹⁰ For an authoritative history of vitalism, see Coleman (1977).

¹¹ Kant was a well-known astronomer as well as a philosopher. He is credited with developing the nebular hypothesis, which explains how planets and stars form from gaseous clouds (nebulae).

Newtonian framework of forces (vital or mechanical); echoing Aristotle, he dubbed living things “organized beings.” Thus, Kant (an avid Newtonian, by the way) brought our understanding of life full circle back to a more refined version of Aristotle’s view.

Kant’s argument was subtle, too subtle for some vitalists, who took it as vindicating their views (Mayr 1997, Ch. 1), and vitalist ideas about life remained popular throughout the nineteenth century and into the early twentieth century; the last eminent biologist with vitalist leanings, Hans Driesch, died in 1941. Other philosophers, for example, J. S. Mill (1843), and scientists, for example, C. Lloyd Morgan (1912), however, recognized that Kant was attacking the foundations of the Newtonian picture of the causal structure of reality.¹² They proposed an alternative to vitalism that has come to be known as emergentism. According to emergentists, there is no special vital substance. Everything (living and nonliving) is made up of the same physical stuff. At certain levels of complexity, however, novel properties and regularities (patterns) arise that cannot be analyzed in terms of the laws of fundamental physics.

As with vitalism, there are different versions of emergentism. The so-called British emergentists were the first to attempt a comprehensive philosophical theory of emergence, detailing its logical structure and how it challenges the picture of causation inherited from classical physics.¹³ A brief overview of British emergentism is therefore instructive. In classical physics, forces act linearly on physical objects and are subject to the principles of vector composition and decomposition. All forces, however complex, are fully decomposable into basic forces (e.g., gravitational and electrical) operating upon pairs of particles of the appropriate kinds (in the cases at hand, mass and charge, respectively). It is here that the British emergentists break with classical physics. The basic particles that make up material objects are collectively organized into increasingly complex structures. Sufficiently complex structures exhibit causal powers that cannot be analyzed as net forces and reduced to those of fundamental physics; the whole is literally greater than the sum of the parts.

According to British emergentism, there is a hierarchy of these causally efficacious, physical structures, beginning, in ascending order, with strictly physical kinds, chemical kinds, biological kinds, and psychological kinds. Each level is

¹² As an avid Newtonian, Kant worried about this and waffled on whether teleological causation represents a fundamental feature of life or whether it represents a mere limitation of the human intellect; he speculated that an alien intelligence might be able to understand life within the framework of classical physics but that humans could not cognize life nonteleologically. It is beyond the scope of this book to pursue the subtleties of Kant’s account; for more on Kant’s views about the teleological character of life, see McLaughlin (1990).

¹³ British emergentism reached its high point in the early twentieth century with the work of philosopher C. D. Broad (1925). For an informative discussion of the history of emergentism, in general, and British emergentism, in particular, see McLaughlin (1992).

characterized by distinctive, causally efficacious, properties that depend upon but cannot be reduced to those occurring at a lower level. As McLaughlin (1992) points out, these (level dependent) novel causal properties amount to “configurational forces.” They can only be exercised by configurations as wholes. Some configurational forces act downwardly on the components of the system, shaping and directing lower-level causal activities. Thus, according to emergentists, complex physical wholes at the biological level (a living antelope) can behave in novel ways (e.g., locomote) in virtue of orchestrating causal interactions among their parts (nerves, muscles, sinews, and bones) that would not otherwise occur. The internally coordinated operation of these lower-level parts critically depends on the organizational integrity of the system as a whole; put starkly, dead antelope, which have lost the pertinent, finely tuned, internal organization, cannot run. The resemblance between configurational forces and Aristotle’s internal regulative principles is striking.

The basic difference between vitalism and emergentism is that the former (in keeping with the theoretical framework of classical physics) seeks to explicate life in terms of novel forces generated by objects bearing special, life conferring, intrinsic properties in a manner analogous to the way in which gravitational force is generated by objects having mass and electrical force is generated by charged objects. Indeed, the association between forces and intrinsic properties of material objects (viewed as composites of basic physical particles) persists to this day. In quantum mechanics, each of the four fundamental forces is carried by a special fundamental particle, that is, a particle having a unique intrinsic causal property: Gluons carry the strong nuclear force, certain bosons the weak nuclear force, photons the force of electromagnetism, and (allegedly) gravitons the force of gravitation.¹⁴ In contrast to vitalism, emergentism holds that living things are constituted by the same kinds of fundamental particles (substance) as inanimate objects. What distinguishes living things from inanimate matter is the manner in which these particles, and the higher-level physical objects (protons, atoms, molecules, etc.) built from them, are collectively organized. When a material system attains the requisite level of organizational complexity it acquires novel functional capacities (e.g., metabolism, locomotion, sensation, and reproduction) that cannot be explicated in terms of lower-level causal interactions among more basic physical parts.

Vitalism and emergentism are sometimes (e.g., Mayr 1997, Ch. 1) grouped with theological accounts as involving “supernatural forces.” This view rests upon the assumption that the only legitimate form of natural causation is ultimately mechanistic or, even more minimalistically, Humean (consisting in nothing more than

¹⁴ One of the central problems of quantum theory is the lack of empirical evidence for gravitons.

the regular succession of events¹⁵). The question of whether all causation is Humean is surely an empirical matter, as opposed to an a priori (logical) truth. As a case in point, quantum phenomena, such as quantum nonlocality, have been held up as challenging the traditional linear concept of causation.¹⁶ For our purposes, however, the important point is that the versions of vitalism and emergentism just discussed, which are among the most fully developed accounts, represent attempts to expand the theoretical framework inherited from classical physics in wholly naturalistic ways to accommodate long standing, seemingly intractable, puzzles about life. The real problem with both vitalism and emergentism is the extremity of the revisions required to flesh them out, namely, the introduction of primitive vital forces and their carriers (vital properties or particles) or alternatively novel forms of causation (primitive configurational forces).

1.4 Darwin to the Rescue?

With the advent of Darwin's theory of evolution by natural selection in the mid-nineteenth century (Darwin 1859), many biologists and philosophers believed that they had found a way of explaining the ostensibly purposive (teleological or goal-directed) characteristics of life in terms of causal mechanisms familiar from classical physics; as Darwin's friend Thomas Huxley gloated, "... Teleology, as commonly understood, ha[s] received its deathblow at Darwin's hands" (Huxley 1864, 82). To revisit a famous example of Darwin's, finches on the Galápagos Islands have beaks of widely differing sizes and shapes that are highly adapted to their food sources. Those eating seeds off the ground, for instance, have shorter, stouter beaks, whereas those feeding on insects have sharper, more slender beaks. It is as if their beaks were deliberately designed to enhance their survival by

¹⁵ The modern scientific concept of causation as *de facto* regularity was developed by the philosopher David Hume (1888) in the eighteenth century. Hume provided powerful, most importantly, empirically based, arguments against the widespread view that causation involves a mysterious, imperceptible, physical connection between causes and their effects. He argued that the relationship of cause to effect is just a matter of invariable succession (time-ordered correlation) – of events of one kind (causes, such as kicking a stone) always being followed by events of another kind (effects, such as the motion of a kicked stone). Hume's analysis of causation continues to dominate thought about causation in natural science (albeit mostly tacitly) and much of philosophy as well.

¹⁶ Quantum nonlocality seems to involve (non-Humean) causes that do not linearly precede their effects but rather occur simultaneously with them, even when vast distances are involved; it is as if the cause "acts" at a distance rather than first being transmitted through an interval of space lying between cause and effect. While quantum nonlocality does not require internal regulative causal principles of the sort being entertained by vitalists and (biologically oriented) emergentists, it does (like them) violate the traditional mechanistic and Humean notion of causation as (ultimately) consisting in linear sequences of causes *followed* by their effects. Indeed, the disquiet of many theoretical physicists and philosophers of physics over quantum phenomena, as underscored by efforts to shore up the "hidden variable" interpretation of quantum mechanics, underscores just how influential the traditional Humean account of causation remains. Indeed, some philosophers and physicists argue that the success of quantum mechanics shows that there is no causation at the quantum level. In any case, however, it is important to keep in mind that quantum nonlocality and teleology provide very different sorts of challenges to the Humean notion of causation.

making it easier to exploit the food sources available to them. But how could nature, left to its own devices, manage something like this?

Aristotle solved this problem by introducing a natural form of goal-directed self-causation: "... it is both by nature and for an end that the swallow makes its nest and the spider its web, and that plants grow leaves for the sake of the fruit and send their roots down (not up) for the sake of nourishment . . ." (*Physics* Bk II 199^{a25}; trans. Barnes 1984, 340). For Aristotle, goal-directed causation is hierarchical. The parts (organs) of organisms have ends (the survival and maintenance of the individual) and organisms, in turn, are parts of eternal species, and hence have higher order ends, namely, the survival and maintenance of the species (Vella 2008, Ch. 3). Many of Darwin's scientific predecessors, eschewing a form of natural causation incompatible with classical physics, attributed the remarkable adaptedness of organisms to their environments as powerful evidence for the existence of God; God deliberately designed them so that they could survive and reproduce in the environments in which he placed them. Darwin's theory of evolution by natural selection advanced a naturalistic explanation for the adaptations of organisms that, unlike neo-Aristotelian vitalism or emergentism, did not presuppose teleological causation.

The core ingredients in Darwin's theory of evolution are a physical source of heritable variation and a natural mechanism for biological change, viz., natural selection. Small heritable variations are always present in a population, providing a potential source of new adaptations. If the environment changes in such a way as to render a heritable characteristic advantageous to the survival and reproduction of an organism, the characteristic will *tend* to become more common in the population; alternatively if it is disadvantageous, it will *tend* to become less common.¹⁷ In other words, the natural environment exerts a selective pressure for advantageous traits and against disadvantageous traits in a manner analogous to the way in which human breeders artificially select for desired traits in plants and animals. It is thus likely that organisms having advantageous traits will eventually come to dominate a population, leading to new varieties and sometimes even new species. The important point, for our purposes, is that there is nothing special about the causation involved. It is admittedly very complex, involving heritable modifications to individual organisms and many interactions of various sorts between organisms and their environments over long periods of time. As the great twentieth

¹⁷ These "tendencies" concern what is statistically expected to happen given a particular *type* of population and a certain *kind* of environment, not what *actually* happens to an individual organism or particular population. An organism with an advantageous trait may fail to survive and reproduce because it has bad luck, suffers an injury, or has some other trait that is very disadvantageous. A pandemic may destroy a whole population of organisms. Darwinian accounts of biological change are thus idealized to the extent that they abstract away from such factors – factors that influence what actually happens to individual organisms and particular populations but are nonetheless extraneous to the evolutionary outcomes (fitness differences) of interest.

century biologist Ernst Mayr (1992) counseled: “Natural selection deals with the properties of individuals of a given generation; it simply does not have any long range goal, even though this may seem so when one looks backward over a long series of generations” (p. 133).

Many of Darwin’s contemporaries were nonetheless convinced that biological evolution is inherently teleological. Even Darwin himself sometimes seems to endorse this view (Greene and Depew 2004, Ch. 7). The challenge for early Darwinians was explaining how an undirected, externally (environmentally) mediated, process of natural selection operating on small hereditary variations (that are random with respect to their adaptive potential) among individual organisms could explain evolutionary trends such as those exhibited in the fossil record. As a result, vitalism and emergentism continued to attract followers into the early years of the twentieth century. A related view, orthogenesis, emerged in the late nineteenth century, and also attracted followers into the early twentieth century. Orthogenesis holds that there are trends in evolution that have no adaptive significance (and hence are difficult to explain in terms of natural selection). It explains these trends by postulating an intrinsic drive (Shanahan 2004). By the middle of the twentieth century, however, teleology was being rapidly expunged from biology as a result of two critical scientific advances: The development of the Modern Evolutionary Synthesis in the 1930s–1940s and the invention of the programmable computer in the 1940s.

The Modern Evolutionary Synthesis unified ideas from a number of biological specialties, including Mendelian genetics, population genetics, embryology, ecology, and paleontology, into a comprehensive, modern evolutionary theory.¹⁸ Heritable variation, upon which natural selection operates, was fleshed out genetically in terms of causally unproblematic notions, viz., mutations that are random (versus directed) with respect to adaptation, recombination, and genetic drift, and the physical basis of heredity, which had evaded Darwin and his contemporaries but was pursued chemically, eventually culminating in the discovery of DNA (Sapp 2003, Ch. 12). Advances in ecology and paleontology, coupled with studies in population genetics, provided scientifically compelling accounts of how natural selection operating on small genetic variations among organisms in natural environments could produce new adaptations and eventually new species over geologically plausible time scales (Greene and Depew 2004, Ch. 9).

With the advent of the programmable computer it became easier to understand how genetic information stored chemically in chromosomes could direct the development of inherited phenotypic (morphological, metabolic, and even behavioral) characteristics of individual organisms. The influence of computer science on

¹⁸ See Huxley (1942) for the now classic statement of the theory.

the development of molecular genetics was profound (Jacob 1973). An organism's genome was likened to a computer program, shaped by natural selection and inherited from its parents. This "genetic program" literally "encodes" all the information required for an organism to construct itself and to continuously regenerate (maintain and repair) itself over the course of its life. The development of the computer metaphor, in the context of rapid advances in molecular biology, left little doubt in the minds of biologists, paleontologists, and philosophers of biology that the ostensibly goal-directed characteristics of life that so impressed Aristotle, and perplexed his eighteenth and nineteenth century successors, could be explicated without presupposing a novel form of causation.

Fleshed out by the Modern Evolutionary Synthesis and the computer metaphor, Darwin's theory does an impressive job of explaining the unity and diversity of known life on Earth. But as discussed below, the fact that it can successfully explain otherwise puzzling relationships among organisms on Earth does not establish that the ability to evolve in this manner is universal let alone essential to life. In addition, Darwin's theory says nothing about how life arises from abiotic chemicals under natural conditions. For, as the title of his *magnum opus* (*On the Origin of Species*) underscores, Darwin's theory *presupposes* an advanced form of life already divided into biological taxa (species) and capable of adaptive evolution.¹⁹ Put briefly, although his contributions to our understanding of life *on Earth* are truly monumental, it is not clear that Darwin is the long-sought Newton of biology.

1.5 Here We Go Again: Aristotelian Roots of Contemporary Accounts of the Nature and Origin(s) of Life

The influence of Aristotle's ideas on contemporary biological thought is most apparent in attempts to answer fundamental questions about life, especially: What is life? How does life arise from nonliving chemicals under natural conditions? Contemporary theories of the nature of life, which are usually articulated as definitions,²⁰ are almost always founded upon characteristic O or R, with one being advanced as more fundamental than the other. Similarly, theories of the origins of life tend to divide along the same lines as definitions of life with regard to these characteristics. As discussed below, however, it is not at all clear that either O or R provides a scientifically fruitful, theoretical foundation for exploring

¹⁹ Creationists who criticize Darwinian evolution because it cannot explain the origins of life either fail to understand this or deliberately suppress it.

²⁰ As discussed in Chapter 4, definitions and scientific theories are different sorts of things; good definitions supply necessary and sufficient conditions whereas good scientific theories are much more open ended. For our purposes here, however, this issue is irrelevant.

deep-seated questions about life. Moreover, even supposing that they do, the perennial problem of goal-directed self-causation remains, for both O and R seem to presuppose it.

1.5.1 Metabolism-Based Versus Evolution-Based Definitions of the Nature of Life

The focus of this section is on the Aristotelian roots of the most influential definitions of life. Most contemporary definitions of life are either metabolism based (characteristic O) or evolution based (characteristic R). Some combine both characteristics in a composite definition, for example, Dyson (1999). Following Aristotle, however, there is a tendency to treat one of these characteristics as more fundamental to life than the other; see the following chapter (Section 2.2) for more detail on these definitions.

Metabolism-based definitions characterize life as a self-organizing system having the capacity to sustain itself against degrading processes for an extended period of time by extracting material, energy or information from its environment. The most popular metabolic definitions can be classified as chemical-metabolic, thermodynamic, or autopoietic. These categories correspond to different levels of abstraction, with chemical-metabolic definitions being the most closely tailored to the metabolic processes of familiar Earth life and autopoietic definitions being the furthest removed. On a chemically based metabolic definition, life is a chemical reaction system that sustains itself by extracting and transforming chemical energy from its environment (e.g., Gánti 2003; Kauffman 1993). Most versions do not require that life uses the same biomolecules and metabolic pathways as familiar life. Some restrict life to organic chemicals. Other versions are open to the possibility of life based on inorganic molecules, for instance, silicon instead of carbon. All chemical-metabolic definitions, however, treat life as a fundamentally chemical phenomenon.

Thermodynamic definitions of life are more abstract. The emphasis is on the physics (versus chemical composition) of the system, namely, its capacity for maintaining bounded regions of local order by extracting energy from thermodynamic gradients found in the environment. Stuart Kauffman (1995, 2000) characterizes such systems as collectively autocatalytic, chemical reaction networks that are far from equilibrium, and argues that they cannot be understood without invoking a new law of thermodynamics. I will have more to say about this shortly.

Chemical-metabolic and thermodynamic definitions differ as to whether life is essentially a chemical or thermodynamic phenomenon. Autopoietic definitions, on the other hand, abstract away from physical as well as chemical detail.

The emphasis is on the “logic” of self-organization, that is, on how the constitutive parts of a system are structurally organized and integrated into a self-sustaining unit. The transition from a nonliving system to a living system is characterized as involving a form of “self-reference” (self-generation, self-maintenance, etc.) that cannot be reduced to relations among its components (e.g., Luisi 1993; Maturana and Varela 1973). Because the focus is on abstract organizational features of a system, autopoietic definitions are sometimes classified as “computational” (Popa 2004, 189). Any system, abstract or concrete, exhibiting the pertinent self-referential, organizational characteristics is (by definition) a living entity. Autopoietic definitions have been used to justify some very radical claims – that ecosystems, human institutions (e.g., corporations), and even Earth itself are individual living things (Lovelock and Margulis 1974; Margulis and Sagan 1995).

Evolutionary definitions of life also come in varying degrees of abstractness. Chemical-evolutionary definitions tend to be Darwinian. Most of them restrict life to nucleic acid-like organic molecules.²¹ NASA’s “chemical Darwinian definition,” which holds that life is a “self-sustaining chemical system capable of Darwinian evolution” (Joyce 1994, xi–xii), is the most widely cited of these definitions. Some evolutionary definitions, however, treat life as a more general statistical-thermodynamic (versus chemical) phenomenon, for example, England (2013) and Goldenfeld and Woese (2011). The most abstract evolutionary definitions conceive of life as purely informational (versus physical or chemical), for example Korzeniewski (2001), Langton (1989), and Trifonov (2011). Darwinian evolution is analyzed as a purely logical process (in essence, a complex algorithm) which in principle could be realized by many different types of systems, all of which *ipso facto* qualify as living things. Evolutionary definitions of this sort resemble autopoietic definitions in abstracting away from chemical and physical detail, but as Section 2.2 discusses, Darwinian evolution is easier to capture in an algorithm than autopoiesis. Algorithmic versions of Darwinian evolution provide the main impetus for the more radical claims of artificial life (ALife) researchers, namely, that they have or soon will create living “creatures” in the information structures of computers; see Section 7.2. The most widely accepted mechanism of evolution is natural selection, but more general mechanisms – for example, Bedau’s (1998) “supple adaptation,” which encompasses not only natural selection but also, for example, Lamarckian selection – are sometimes invoked in evolutionary definitions of life.

²¹ Following Schrödinger (1944), who famously anticipated the structure of DNA when he argued that hereditary information must be encoded on an “aperiodic solid” (p. 5), it is widely thought that all chemical genetic systems will consist of aperiodic polymers, and that (due to some unusual chemical constraints) molecules having a nucleic acid-like structure are the only ones capable of supporting evolution by natural selection (Benner and Hutter 2002).

The concept of metabolism underlying metabolic definitions of life is strikingly similar to Aristotle's notion of self-nutrition. Metabolic definitions emphasize the capacity of a living system to sustain *itself* as an "organized being" by extracting and processing "nutrients" (material, energetic, or even purely informational) found in its environment. The influence of Aristotle's ideas on evolutionary definitions is more tenuous. Both Darwin and Aristotle take reproduction to be essential to life. However, their account of the relation between reproduction, a process undergone by individual organisms, and similarities and differences among groups of organisms, such as species, differs. Writing two thousand years before Darwin, Aristotle believed that species are eternal. He focused on the role of reproduction in producing descendants closely resembling their progenitors. As part of an eternal species, the individual organism has an internal, goal-directed, drive to reproduce so that it can sustain the species to which it belongs. Contemporary evolution-based definitions of life, in contrast, attribute similarities and differences among groups of organisms to an undirected, externally mediated, process of differential survival and reproduction; organisms best adapted to a given environment are more likely to survive and leave descendants than those that are less well adapted.

Individual organisms are the paradigmatic living things. Evolution-based definitions of life, however, have trouble classifying them as such. Organisms do not evolve as individuals; they merely live and die. Some proponents of evolutionary definitions bite the bullet and simply deny that individual organisms qualify as living things. Others (e.g., Bedau 1998; Benner 2010) treat them as subordinate cases of life; see Section 2.2.3 for a discussion. Metabolism-based definitions, in contrast, admit the possibility of living things that are incapable of reproducing. For our purposes, the important point is that metabolism-based definitions of life construe the individual organism as the basic unit of life whereas evolution-based definitions take groups of related organisms (populations, species, or lineages) as the basic unit of life. This represents a fundamental difference in the theoretical frameworks underlying metabolic and evolutionary definitions of life – a difference presaged by the old Aristotelian controversy over whether nutrition (maintenance of the individual) or reproduction (continuity of the species) is more essential to life.

As we have seen, there is a long tradition of framing the nature of life in terms of either characteristic O or R, but not both. In truth, however, there is little contemporary scientific support for this rivalry. With the exception of viruses, whose status as living is controversial, all known life exhibits both characteristics. Moreover, in familiar life, these functions are deeply entangled at the molecular level. The biomolecules realizing characteristics O and R are different in type: Proteins supply the bulk of the structural and catalytic material required for

building, maintaining, and repairing bodies, whereas a very different molecular species, nucleic acids, stores (DNA) and manages (RNA) the hereditary material required for genetic-based reproduction. Minuscule, very complex molecular machines (ribosomes), composed of both RNA and protein, which are found in large numbers in every cell, coordinate these functions. The information stored in DNA is transcribed onto RNA and subsequently translated into proteins by ribosomes. This complex, highly organized, molecular architecture provides the physical basis for metabolism, development, reproduction, and evolution. In other words, from a molecular perspective, there is little reason to suppose that characteristic R is more fundamental than O, or vice versa.

But even supposing that O and/or R were universal to known life, it does not follow that they are fundamental to life. They could be symptoms of more basic, but as yet unknown, properties of life. As an analogy, consider water. Various perceptible characteristics thought to be universal to water (e.g., being wet, transparent, and a good solvent) were traditionally used to distinguish it from other chemical substances. But this information does not reveal the underlying nature of water, namely, that it is composed primarily of molecules of H₂O. Moreover, the latter information cannot even be inferred from its triple point, a distinctive combination of pressure and temperature at which its gas, liquid, and solid phases coexist in equilibrium, which does uniquely distinguish water from all other (pure) chemical substances. The point is the modern scientific account of water required the development of a new theoretical framework for understanding chemical substance. The old Aristotelian concept of chemical substance, which dominated scientific thought until the pioneering work of Lavoisier, in the late eighteenth century, held that water is a basic (indivisible) chemical element. Lavoisier (1783) challenged this view by presenting compelling empirical evidence that water is a chemical compound, paving the way for the development of molecular theory.

Viewed in light of the above discussion, one cannot help but wonder whether characteristics O and R could be misleading us about the nature of life. That is, are characteristics that are merely symptomatic of life being elevated to the status of essential characteristics because they are thought to be universal to life? The situation is actually a bit more worrisome. For as Chapter 5 discusses, there is compelling empirical and theoretical evidence that characteristics O and R may not only be misleading but also are unreliable for discriminating living from nonliving systems. It could turn out that O and R are neither necessary nor sufficient for life.

1.5.2 Metabolism-First Versus Genes-First Theories of the Origin(s) of Life

Scientific theories of the origin(s) of life typically bifurcate along the same lines as definitions of life into genes-first models, which focus on the capacity to reproduce

and undergo evolution by natural selection (characteristic R), and metabolism-first models, which focus on the capacity to self-organize and maintain self-organization for an extended period of time (characteristic O).²² The most influential versions are the RNA World, which is a genes-first theory, and the Small Molecule (SM) World, which is a metabolism-first theory.²³ As in the previous section, our concern here is with the Aristotelian roots of these theories; they are discussed in much greater detail in Section 5.4.

Proteins are the primary catalysts used by the metabolic processes of familiar life. They consist of long chains (polymers) of amino acids. According to the SM World,²⁴ life originates with the development of a sufficiently complex, collectively autocatalytic, chemical reaction system involving small molecules such as amino acids, cofactors, and peptides (short chains of as few as two amino acids). But while abiotic amino acids are found in the natural environment, abiotic peptides are not. The chemical reactions required for forming peptide bonds between amino acids are thermodynamically uphill and difficult to envision occurring abiotically in a natural (versus controlled laboratory) environment. This is not the most serious challenge facing the SM World, however. The question of how a minimally autocatalytic chemical reaction network capable of increasing in complexity to the point of qualifying as proto-metabolic could arise *de novo* in an open thermodynamic system consisting of a variety of small molecules, including peptides, is not fleshed out.

An incipient autocatalytic chemical reaction network capable of developing the organizational complexity of a proto-metabolism must be able to sustain itself in the face of spontaneously degrading (thermodynamically downhill) geochemical processes.²⁵ The probability of a reaction network of this sort arising under natural conditions from a chemically heterogeneous and unorganized reaction system of small molecules is (given the current state of our knowledge) very small (Orgel 2008). Proponents of the SM World tend to gloss over this difficulty by appealing to an unexplained process of “spontaneous chemical self-organization.” Instead of solving the problem, however, this amounts to resurrecting the old Aristotelian

²² As discussed in Section 5.4.2, there is a long (and mistaken) history of thinking that one can infer an account of the origin of life from an account of the nature of life.

²³ Like the definitions of life discussed earlier, the SM World and RNA World models encompass a variety of more specific versions, differing in chemical and physical details. Chapter 5 discusses how the most popular versions of these respective theories differ among each other.

²⁴ Shapiro (2006) coined the term but the concept goes back to Oparin (1964) and has been developed along different lines by de Duve (1995), Dyson (1999), Kauffman (1993), Shapiro (2006), and Wächtershäuser (1990), among others.

²⁵ As Chapter 5 discusses, some mechanism for excluding certain molecules while admitting and concentrating others is needed. Various possibilities have been suggested, e.g., membranous compartments (e.g., micelles, aerosols, and lipid vesicles), absorbent mineral surfaces, and porous rocks, but this just pushes the problem back a step insofar as it raises the issue of how a mechanism able to achieve the sophisticated level of molecular selectivity and concentration required could arise under natural conditions; see Cleland (2013) for a more detailed discussion.

idea of goal-directed self-causation in a scientifically less objectionable form. Stuart Kauffman (e.g., 1995, 2000) is one of the few researchers who fully grasp this point. According to Kauffman, an understanding of the origin of life requires recourse to new thermodynamic principles of self-organization in complex, far from equilibrium reaction systems. When thermodynamically open reaction systems attain a certain level of complexity, they spontaneously self-organize into collectively autocatalytic reaction networks capable of sustaining themselves against internal and external perturbations and increasing their organizational complexity.

There is an intriguing parallel between Kauffman's ideas and those of the British emergentists. Indeed, Kauffman and fellow scientific travelers (e.g., Hazen 2005) routinely use the term "emergence" to characterize the processes involved in the transition from chemistry to biology. Both approaches propose specific modifications to physical theory, in the case of the British empiricists, Newtonian principles of vector analysis, and in the case of Kauffman, the foundations of statistical thermodynamics. Such a strategy has the advantage of suggesting definite ways – candidates for further scientific investigation – in which our theoretical understanding of life as a natural phenomenon may be inadequate. For our purposes, however, the important point is that both approaches are neo-Aristotelian in character. They distinguish living from inanimate material systems by invoking a natural form of goal-directed self-causation that cannot be analyzed in terms of causal mechanisms familiar to physical scientists.

In contrast to the SM World, the RNA World holds that life originated with the development of a proto-genetic system from a "prebiotic soup" of small RNA molecules and their molecular precursors.²⁶ The impetus for the RNA World is the Nobel Prize winning discovery that (unlike DNA) some RNA molecules are self-replicating. This discovery is held up as solving a formidable "[which came first] the chicken or egg" dilemma: The replication of nucleic acids (DNA and RNA) was previously thought to require protein enzymes and the synthesis of protein enzymes depends upon nucleic acids (Orgel 2004).

The abiotic synthesis of a small RNA molecule (RNA oligomer) under natural conditions is even more challenging, however, than the synthesis of a short peptide. An RNA molecule consists of a long chain of ribonucleotides. A ribonucleotide is made up of three molecular subunits, a phosphate unit bonded to a sugar (ribose) unit bonded to a nucleobase (purine or pyrimidine) in a precisely organized, three-dimensional pattern. The assembly of an RNA oligomer in a laboratory setting requires a long sequence of carefully staged and controlled

²⁶ The term "RNA World" was coined by Gilbert (1986) but the concept was developed earlier by Woese (1967, 179–195), Crick (1968), and Orgel (1968).

chemical reactions. Given our current understanding of geochemistry, the probability of an RNA oligomer being abiotically synthesized under natural conditions is extremely low (e.g., Shapiro 2000, 2006). Moreover, the likelihood that such a molecule would be self-catalytic is even lower; most RNA molecules are not self-replicating. Indeed, the theoretical difficulties in making good chemical sense of the synthesis of self-replicating RNA under natural conditions are so serious that many researchers speculate that the RNA World must have been preceded by an earlier “Pre-RNA World” based on a different self-replicating molecule (e.g., Joyce and Orgel 1999). Various candidates have been suggested but, as discussed in Section 5.4, none seems much easier to synthesize than RNA.

The central focus of the RNA World is on “evolving” a population of self-replicating RNA molecules with increasingly efficient and versatile catalytic capacities. Most scenarios begin with a primordial pool of mutually self-catalytic RNA oligomers. In order for the RNA World to even get off the ground, however, the pool of molecules must have a fairly sophisticated level of chemical organization. A pool of RNA oligomers arising under natural (versus artificial) conditions would include a variety of other chemical species. An incipient RNA World would thus be subject to degrading chemical reactions unless there was some mechanism for excluding them.²⁷ But even supposing that it consisted exclusively of RNA oligomers, an embryonic RNA World would be subject to cross-reactions with nonreplicating RNA. In short, one needs just the right mix of just the right types of RNA oligomers in a reasonably clean environment for the system to be capable of evolving in the manner required. Proponents of the RNA World tend to downplay this problem by appealing to the “spontaneous emergence” of self-organization in mutually catalytic populations of small RNA oligomers. As in the case of the SM World, however, downplaying the problem does not solve it, but instead tacitly resurrects the concept of goal-directed self-causation in scientifically more palatable language.

In addition to the problems just discussed, both the RNA World and the SM World have difficulty explaining the origins of the deep molecular entanglement between the metabolic and genetic systems of familiar Earth life. This difficulty is not surprising. By positioning themselves as rivals with respect to this high level architectural division, they in effect marginalize the issue of how these systems became chemically integrated; their interdependence falls into the crack between the theories. Defenders of the SM World conjecture that nucleotides and RNA oligomers arose in a developing autocatalytic reaction network to serve some

²⁷ As in the case of the SM World, various natural mechanisms for concentrating and protecting delicate biomolecules, in this case RNA oligomers and their building blocks, have been suggested, namely, membranous compartments of various sorts, absorbent mineral surfaces and porous rocks. These and other possibilities are explored Chapter 5.

nongenetic purpose and took on a genetic role much later, after the development of a primitive metabolism (e.g., Shapiro 2006). Defenders of the RNA World, on the other hand, contend that RNA molecules performed both genetic and primitive metabolic functions until some time after the appearance of the first “riboorganisms” (Benner and Ellington 1987). At some point RNA began synthesizing peptides and proteins, which being better catalysts, eventually took over the metabolic functions of RNA. Given the size and complexity of proteins and nucleic acids, however, it seems unlikely that the complex cooperative arrangement between them arose at such a late stage in the development of life. A few researchers have developed hybrid theories in an effort to deal with this problem. Copley et al. (2005), for instance, argue that amino acids became associated with the nucleobases that code for them in a small molecule reaction network much earlier, prior to the synthesis of either proteins or nucleic acids. For our purposes, however, the important point is that the ancient architectural division of life along the lines of characteristics O and R seems to be driving theorizing about the origin of life.

1.6 Concluding Thoughts

To wrap up, contemporary scientific and philosophical thought about the nature and origin(s) of life is founded upon the supposition that what distinguishes life from nonlife is the capacity of a system to (1) self-organize and maintain self-organization for an extended period of time (O) and (2) undergo genetic-based reproduction (R). Precursors to O and R are found in the ancient writings of Aristotle, who identified “nutrition” and “reproduction” as the most basic characteristics of life and debated which one is more fundamental. Following Aristotle, most contemporary scientists and philosophers privilege one of these characteristics as more basic to life than the other. But even supposing, for the sake of argument, that these neo-Aristotelian characteristics are universal to life, it does not follow that they are fundamental to life. As we have seen, they could be symptoms of more basic but as yet unknown properties of life, in which case it would be a mistake to try to found a universal theory of life upon them. Likewise, most theories of the origin(s) of life are based upon variations on O or R (sometimes both), further underscoring the influence of Aristotle’s ideas on modern attempts to answer fundamental questions about life. In short, the contemporary quest for answers to foundational biological questions is deeply rooted in Aristotelian assumptions. As Chapter 4 discusses, the Aristotelian character of biological thought may in part explain why we lack a truly general theory for even familiar life (let alone life considered generally); when presented with an allegedly general principle for life, many biologists delight in listing counterexamples.

As emphasized in this chapter, the most challenging aspect of characteristics O and R is making good sense of them causally. Unlike most contemporary scientific theories about natural phenomena, O and R characterize life functionally. Functional characterizations are highly abstract, which helps to explain why some ALife researchers are willing to countenance purely informational life implemented on digital computers; I will have more to say about computer life in Chapter 7. Functional properties are peculiar insofar as they are instantiated by an entity *before* the ends that functionally define them actually occur. That is, they are intrinsically goal directed. The concept of intrinsically goal-directed processes represents a problem for contemporary scientists and philosophers of science whose concept of causation holds that causes precede their effects without anticipating them. For this reason, functional capacities are rarely if ever taken as basic in science. Instead, they are treated as derivative, that is, as products of familiar, undirected, causal interactions among the components of systems exhibiting them. Mental states, which are *prima facie* intrinsically goal directed (e.g., my desiring a candy bar), provide an especially good illustration. Psychology treats them as functional properties. But even those who defend the autonomy of psychology – its irreducibility to neuroscience (and ultimately to chemistry and physics) – are loath to interpret mental states as requiring a truly novel form of physical causation. In the words of philosopher Jonathan Lowe (1993), “[t]hat mental phenomena are part of the natural, causal order of events is surely not to be denied” (p. 629). As with the case of life, the scientific challenge is showing how mental properties are generated by physicochemical processes through unproblematic causal interactions.

As discussed in Section 1.4, Darwin’s theory of evolution by natural selection is commonly held up as showing how the ostensibly goal-directed characteristics of life can be explained in terms of linear chains of undirected causes and effects. The remarkable adaptedness of organisms to their environment is explicated as the product of a blind process of natural selection operating on heritable variations (mutations, recombination, and genetic drift) over long periods of time. The causal processes involved are complex, involving an enormous number of diverse interactions between organisms and their environments, but the causation involved is Humean. As discussed, however, even when fleshed out in light of twentieth century developments in biology, Darwin’s theory cannot account for the transition from chemistry to biology. Admittedly, Darwin speculated (in an 1870 letter to his good friend the famous botanist Joseph Hooker) about life originating in a “warm little pond.” But he says nothing about the nature of the causal processes involved.

Many advocates of the RNA World follow Manfred Eigen (1992) in invoking a principle of *molecular* natural selection to explain the development of the earliest living chemical systems from a population of (depending upon the version) self or

collectively autocatalytic RNA oligomers. Primordial, self-catalytic RNA oligomers are characterized as “encoding” their catalytic efficiency as hereditary information in a neo-Darwinian sense (e.g., Alberts et al. 2002, Ch. 6). But the concept of encoding being invoked is so attenuated from a biological perspective that it is not clear that it can do the work required of it. A true encoding device converts information from one form to another. In the RNA World, however, there is no distinction between the encoding molecule and the encoded molecule; in the simplest version of the RNA World they are one and the same. In classical Darwinian evolution, in contrast, the distinction between the phenotype (metabolic, structural, and behavior characteristics) and the genotype of an organism is central to the selection process; the environment acts directly on the encoded phenotype and only indirectly on the genotype. In short, it is not clear that the specter of goal-directed causation can be expunged from the RNA World in a manner analogous to the way in which Darwin expunged it from advanced life.

But even supposing, for the sake of argument, that one could successfully explain the “evolution” of a primitive “riboorganism” (Benner 1987, 53) from a primordial collection of RNA oligomers in terms of a process of molecular natural selection, there is still the problem, discussed in Section 1.5.2, of the prebiotic assembly of the first catalytic RNA oligomers. Appealing to an even more general principle of chemical natural selection, which could just as easily be invoked by advocates of the SM World, will not work. For in the absence of greater specificity about the processes involved, such an appeal amounts to little more than invoking the (in essence, chemically miraculous) “spontaneous” self-assembly of a favored primordial biomolecule. The more sweeping and removed from biology a principle of natural selection, the less light it is able to shed upon the processes actually involved in the origins of life. This is perhaps most easily appreciated when one considers that some physicists have generalized natural selection even further, to a principle of *cosmological* natural selection (e.g., Smolen 1997), in efforts to explain the historical development of the universe. Such principles are too vague and sweeping to have genuine explanatory power for something as specific as the origin or nature of biological life. As Popper (1963) cautioned, the capacity of a scientific theory to explain everything is not strength but rather a weakness.

We are left in a quandary. Was Aristotle correct? Do answers to the most fundamental questions about life – its nature and origins – require recourse to a primitive form of goal-directed self-causation? Or, alternatively, does the difficulty in satisfactorily answering these questions lie in commitment to an unpropitious foundation of basic concepts and principles for theorizing about life? Put more provocatively, do we need to stop constructing theories of the nature and origin(s) of life on the basis of a theoretical framework founded more than two thousand years ago in ancient Greece?

During the last few decades the neo-Aristotelian concept of emergence has resurfaced in scientific and philosophical discussions about life. Whether an intelligible concept of emergence can be developed remains an interesting and open question, well worth pursuing. Some philosophers, most notably Kim (1992, 1999), contend that the concept is logically incoherent. Others (e.g., Bedau 1997; Humphreys 1996; Wimsatt 2007) are actively working on surmounting the logical problems identified by Kim and others. This book takes the road less traveled: It rejects a neo-Aristotelianism theoretical framework for reasoning about life. My choice is guided in part by skepticism about the coherence of the concept of goal-directed self-causation (tacitly appealed to in the concept of emergence) and in part by the history of science. Indeed, as Chapter 4 discusses, the rapid development and greatest empirical successes of the other natural sciences (most notably, chemistry and physics) occurred *after* the abandonment of Aristotelian concepts and principles. It strikes me as a bit scandalous that biology is the only natural science whose foundations are still dominated (albeit often tacitly) by neo-Aristotelian ideas. In any case, it seems clear that exploring alternative theoretical frameworks for theorizing about life is worthwhile, especially in light of some of the remarkable discoveries, discussed in Chapters 5, 6, 7, and 8, being made by molecular biologists and microbiologists.