DNA and the Origin of Life: Information, Specification, and Explanation

Stephen C. Meyer

Theories about the origin of life necessarily presuppose knowledge of the attributes of living cells. As historian of biology Harmke Kamminga has observed, "At the heart of the problem of the origin of life lies a fundamental question: What is it exactly that we are trying to explain the origin of?"[1] Or as the pioneering chemical evolutionary theorist Alexander Oparin put it, "The problem of the nature of life and the problem of its origin have become inseparable."[2] Origin-of-life researchers want to explain the origin of the first and presumably simplest—or, at least, minimally complex—living cell. As a result, developments in fields that explicate the nature of unicellular life have historically defined the questions that origin-of-life scenarios must answer.

Since the late 1950s and 1960s, origin-of-life researchers have increasingly recognized the complex and specific nature of unicellular life and the biomacromolecules on which such systems depend. Further, molecular biologists and origin-of-life researchers have characterized this complexity and specificity in informational terms. Molecular biologists routinely refer to DNA, RNA, and proteins as carriers or repositories of "information."[3] Many origin-of-life researchers now regard the origin of the information in these

biomacromolecules as the central question facing their research. As Bernd-Olaf Kuppers has stated, "The problem of the origin of life is clearly basically equivalent to the problem of the origin of biological information."[4]

This essay will evaluate competing explanations for the origin of the information necessary to build the first living cell. To do so will require determining what biologists have meant by the term *information* as it has been applied to biomacromolecules. As many have noted, "information" can denote several theoretically distinct concepts. This essay will attempt to eliminate this ambiguity and to determine precisely what type of information origin-of-life researchers must explain "the origin of." What follows will first seek to *characterize* the information in DNA, RNA, and proteins as an *explanandum* (a fact in need of explanation) and, second, to *evaluate* the efficacy of competing classes of explanation for the origin of biological information (that is, the competing *explanans*).

Part I will seek to show that molecular biologists have used the term *information* consistently to refer to the joint properties of *complexity* and functional *specificity* or *specification*. Biological usage of the term will be contrasted with its classical information-theoretic usage to show that "biological information" entails a richer sense of information than the classical mathematical theory of Shannon and Wiener. Part I will also argue against attempts to treat biological "information" as a metaphor lacking empirical content and/or ontological status.[5] It will show that the term *biological information* refers to two real features of living systems, complexity and specificity, features that jointly do require explanation.

Part II will evaluate competing types of explanation for the origin of the specified biological information necessary to produce the first living system. The categories of "chance" and "necessity" will provide a helpful heuristic for understanding the recent history of origin-of-life research. From the 1920s to the mid-1960s, origin-of-life researchers relied heavily on theories emphasizing the creative role of random events—"chance"—often in tandem with some form of prebiotic natural selection. Since the late 1960s, theorists have instead emphasized deterministic self-organizational laws or properties—that is, physical-chemical "necessity."

Part II will critique the causal adequacy of chemical evolutionary theories based on "chance," "necessity," and the combination of the two.

A concluding part III will suggest that the phenomenon of information understood as specified complexity requires a radically different explanatory approach. In particular, I will argue that our present knowledge of causal powers suggests intelligent design as a better, more causally

adequate explanation for the origin of the specified complexity (the information so defined) present in large biomolecules such as DNA, RNA, and proteins.

I.

A. Simple to Complex: Defining the Biological *Explanandum*

After Darwin published the *Origin of Species* in 1859, many scientists began to think about a problem that Darwin had not addressed.[6] Although Darwin's theory purported to explain how life could have grown gradually more complex starting from "one or a few simple forms," it did not explain, or attempt to explain, how life had first originated. Yet in the 1870s and 1880s, evolutionary biologists like Ernst Haeckel and Thomas Huxley assumed that devising an explanation for the origin of life would be fairly easy, in large part because Haeckel and Huxley assumed that life was, in its essence, a chemically simple substance called "protoplasm" that could easily be constructed by combining and recombining simple chemicals such as carbon dioxide, oxygen, and nitrogen.

Over the next sixty years, biologists and biochemists gradually revised their view of the nature of life. During the 1860s and 1870s, biologists tended to see the cell, in Haeckel's words, as an undifferentiated and "homogeneous globule of plasm." By the 1930s, however, most biologists had come to see the cell as a complex metabolic system.[7] Origin-of-life theories reflected this increasing appreciation of cellular complexity. Whereas nineteenth-century theories of abiogenesis envisioned life arising almost instantaneously via a one- or two-step process of chemical "autogeny," early twentieth-century theories, such as Oparin's theory of *evolutionary* abiogenesis, envisioned a multibillion-year process of transformation from simple chemicals to a complex metabolic system.[8] Even so, most scientists during the 1920s and 1930s still vastly underestimated the complexity and specificity of the cell and its key functional components—as developments in molecular biology would soon make clear.

B. The Complexity and Specificity of Proteins

During the first half of the twentieth century, biochemists had come to recognize the centrality of proteins to the maintenance of life. Although many mistakenly believed that proteins also contained the source of heredity information, biologists repeatedly underestimated the complexity of proteins.

For example, during the 1930s, English X-ray crystallographer William Astbury elucidated the molecular structure of certain fibrous proteins, such as keratin, the key structural protein in hair and skin.[9] Keratin exhibits a relatively simple, repetitive structure, and Astbury was convinced that all proteins, including the mysterious globular proteins so important to life, represented variations on the same primal and regular pattern. Similarly, biochemists Max Bergmann and Carl Niemann of the Rockefeller Institute argued in 1937 that the amino acids in proteins occurred in regular, mathematically expressible proportions. Other biologists imagined that insulin and hemoglobin proteins, for example, "consisted of bundles of parallel rods."[10]

Beginning in the 1950s, however, a series of discoveries caused this simplistic view of proteins to change. From 1949 to 1955, biochemist Fred Sanger determined the structure of the protein molecule, insulin. Sanger showed that insulin consisted of a long and irregular sequence of the various amino acids, rather like a string of differently colored beads arranged without any discernible pattern. His work showed for a single protein what subsequent work in molecular biology would establish as a norm: The amino acid sequence in functional proteins generally defies expression by any simple rule and is characterized instead by aperiodicity or complexity.[11] Later in the 1950s, work by John Kendrew on the structure of the protein myoglobin showed that proteins also exhibit a surprising three-dimensional complexity. Far from the simple structures that biologists had imagined earlier, an extraordinarily complex and irregular three-dimensional shape was revealed: a twisting, turning, tangle of amino acids. As Kendrew explained in 1958, "The big surprise was that it was so irregular . . . the arrangement seems to be almost totally lacking in the kind of regularity one instinctively anticipates, and it is more complicated than has been predicted by any theory of protein structure."[12]

By the mid-1950s, biochemists recognized that proteins possess another remarkable property. In addition to their complexity, proteins also exhibit specificity, both as one-dimensional arrays and three-dimensional structures. Whereas proteins are built from chemically rather simple amino acid "building blocks," their function (whether as enzymes, signal transducers, or structural components in the cell) depends crucially on a complex but specific arrangement of those building blocks.[13] In particular, the specific sequence of amino acids in a chain and the resultant chemical interactions between amino acids largely determine the specific three-dimensional structure that the chain as a whole will adopt. Those structures or shapes

in turn determine what function, if any, the amino acid chain can perform in the cell.

For a functioning protein, its three-dimensional shape gives it a hand-in-glove fit with other molecules, enabling it to catalyze specific chemical reactions or to build specific structures within the cell. Because of its three-dimensional specificity, one protein can usually no more substitute for another than one tool can substitute for another. A topoisomerase can no more perform the job of a polymerase than a hatchet can perform the function of a soldering iron. Instead, proteins perform functions only by virtue of their three-dimensional specificity of fit, either with other equally specified and complex molecules or with simpler substrates within the cell. Moreover, the three-dimensional specificity derives in large part from the one-dimensional sequence specificity in the arrangement of the amino acids that form proteins. Even slight alterations in sequence often result in the loss of protein function.

C. The Complexity and Sequence Specificity of DNA

During the early part of the twentieth century, researchers also vastly underestimated the complexity (and significance) of nucleic acids such as DNA and RNA. By then, scientists knew the chemical composition of DNA. Biologists and chemists knew that in addition to sugars (and later phosphates), DNA was composed of four different nucleotide bases, called adenine, thymine, cytosine, and guanine. In 1909, chemist P. A. Levene showed (incorrectly as it later turned out) that the four different nucleotide bases always occurred in equal quantities within the DNA molecule.[14] He formulated what he called the "tetranucleotide hypothesis" to account for that putative fact. According to that hypothesis, the four nucleotide bases in DNA linked together in repeating sequences of the same four chemicals in the same sequential order. Since Levene envisioned those sequential arrangements of nucleotides as repetitive and invariant, their potential for expressing any genetic diversity seemed inherently limited. To account for the heritable differences between species, biologists needed to discover some source of variable or irregular specificity, some source of information, within the germ lines of different organisms. Yet insofar as DNA was seen as an uninterestingly repetitive molecule, many biologists assumed that DNA could play little if any role in the transmission of heredity.

That view began to change in the mid-1940s for several reasons. First, Oswald Avery's famous experiments on virulent and nonvirulent strains of *Pneumococcus* identified DNA as the key factor in accounting for heritable differences between different bacterial strains.[15] Second, work by Erwin Chargaff of Columbia University in the late 1940s undermined the "tetranucleotide hypothesis." Chargaff showed, contradicting Levene's earlier work, that nucleotide frequencies actually do differ between species, even if they often hold constant within the same species or within the same organs or tissues of a single organism.[16] More important, Chargaff recognized that even for nucleic acids of exactly "the same analytical composition"—meaning those with the same relative proportions of the four bases (abbreviated A, T, C, and G)—"enormous" numbers of variations in sequence were possible. As he put it, different DNA molecules or parts of DNA molecules might "differ from each other . . . in the sequence, [though] not the proportion, of their constituents." As he realized, for a nucleic acid consisting of 2,500 nucleotides (roughly the length of a long gene) the number of sequences "exhibiting the same molar proportions of individual purines [A, G] and pyrimidines [T, C] . . . is not far from 10^{1500}."[17] Thus, Chargaff showed that, contrary to the tetranucleotide hypothesis, base sequencing in DNA might well display the high degree of variability and aperiodicity required by any potential carrier of heredity.

Third, elucidation of the three-dimensional structure of DNA by Watson and Crick in 1953 made clear that DNA could function as a carrier of hereditary information.[18] The model proposed by Watson and Crick envisioned a double-helix structure to explain the Maltese-cross pattern derived from X-ray crystallographic studies of DNA by Franklin, Wilkins, and Bragg in the early 1950s. According to the now well-known Watson and Crick model, the two strands of the helix were made of sugar and phosphate molecules linked by phosphodiester bonds. Nucleotide bases were linked horizontally to the sugars on each strand of the helix and to a complementary base on the other strand to form an internal "rung" on a twisting "ladder." For geometric reasons, their model required the pairing (across the helix) of adenine with thymine and cytosine with guanine. That complementary pairing helped to explain a significant regularity in composition ratios discovered by Chargaff. Though Chargaff had shown that none of the four nucleotide bases appears with the same frequency as all the other three, he did discover that the molar proportions of adenine and thymine, on the one hand, and cytosine and guanine, on the other, do

consistently equal each other.[19] Watson and Crick's model explained the regularity Chargaff had expressed in his famous "ratios."

The Watson-Crick model made clear that DNA might possess an impressive chemical and structural complexity. The double-helix structure for DNA presupposed an extremely long and high-molecular-weight structure, possessing an impressive potential for variability and complexity in sequence. As Watson and Crick explained, "The sugar-phosphate backbone in our model is completely regular but any sequence of base pairs can fit into the structure. It follows that in a long molecule many different permutations are possible, and it, therefore, seems likely that the precise sequence of bases is the code which carries genetic information."[20]

As with proteins, subsequent discoveries soon showed that DNA sequences were not only complex but also highly specific relative to the requirements of biological function. Discovery of the complexity and specificity of proteins had led researchers to suspect a functionally specific role for DNA. Molecular biologists, working in the wake of Sanger's results, assumed that proteins were much too complex (and yet also functionally specific) to arise by chance in vivo. Moreover, given their irregularity, it seemed unlikely that a general chemical law or regularity could explain their assembly. Instead, as Jacques Monod has recalled, molecular biologists began to look for some source of information or "specificity" within the cell that could direct the construction of such highly specific and complex structures. To explain the presence of the specificity and complexity in the protein, as Monod would later insist, "you absolutely needed a code."[21]

The structure of DNA as elucidated by Watson and Crick suggested a means by which information or "specificity" might be encoded along the spine of DNA's sugar-phosphate backbone.[22] Their model suggested that variations in sequence of the nucleotide bases might find expression in the sequence of the amino acids that form proteins. In 1955, Crick proposed this idea as the so-called sequence hypothesis. According to Crick's hypothesis, the specificity of arrangement of amino acids in proteins derives from the specific arrangement of the nucleotide bases on the DNA molecule.[23] The sequence hypothesis suggested that the nucleotide bases in DNA functioned like letters in an alphabet or characters in a machine code. Just as alphabetic letters in a written language may perform a communication function depending on their sequence, so, too, might the nucleotide bases in DNA result in the production of a functional protein molecule depending on their precise sequential arrangement. In both cases, function

depends crucially on sequence. The sequence hypothesis implied not only the complexity but also the functional specificity of DNA base sequences.

By the early 1960s, a series of experiments had confirmed that DNA base sequences play a critical role in determining amino acid sequence during protein synthesis.[24] By that time, the processes and mechanisms by which DNA sequences determine key stages of the process were known (at least in outline). Protein synthesis or "gene expression" proceeds as long chains of nucleotide bases are first copied during a process known as transcription. The resulting copy, a "transcript" made of single-stranded "messenger RNA," now contains a sequence of RNA bases precisely reflecting the sequence of bases on the original DNA strand. The transcript is then transported to a complex organelle called a ribosome. At the ribosome, the transcript is "translated" with the aid of highly specific adaptor molecules (called transfer-RNAs) and specific enzymes (called amino-acyl tRNA synthetases) to produce a growing amino acid chain (figure 1).[25] Whereas the function of the protein molecule derives from the specific arrangement of twenty different types of amino acids, the function of DNA depends on the arrangement of just four kinds of bases. This lack of a one-to-one correspondence means that a group of three DNA nucleotides (a triplet) is needed to specify a single amino acid. In any case, the sequential arrangement of the nucleotide bases determines (in large part) the one-dimensional sequential arrangement of amino acids during protein synthesis.[26] Since protein function depends critically on amino acid sequence and amino acid sequence depends critically on DNA base sequence, the sequences in the coding regions of DNA themselves possess a high degree of specificity relative to the requirements of protein (and cellular) function.

D. Information Theory and Molecular Biology

From the beginning of the molecular biological revolution, biologists have ascribed information-bearing properties to DNA, RNA, and proteins. In the parlance of molecular biology, DNA base sequences contain the "genetic information" or the "assembly instructions" necessary to direct protein synthesis. Yet the term *information* can denote several theoretically distinct concepts. Thus, one must ask which sense of "information" applies to these large biomacromolecules. We shall see that molecular biologists employ both a stronger conception of information than that of mathematicians and information-theorists and a slightly weaker conception of the term than that of linguists and ordinary users.

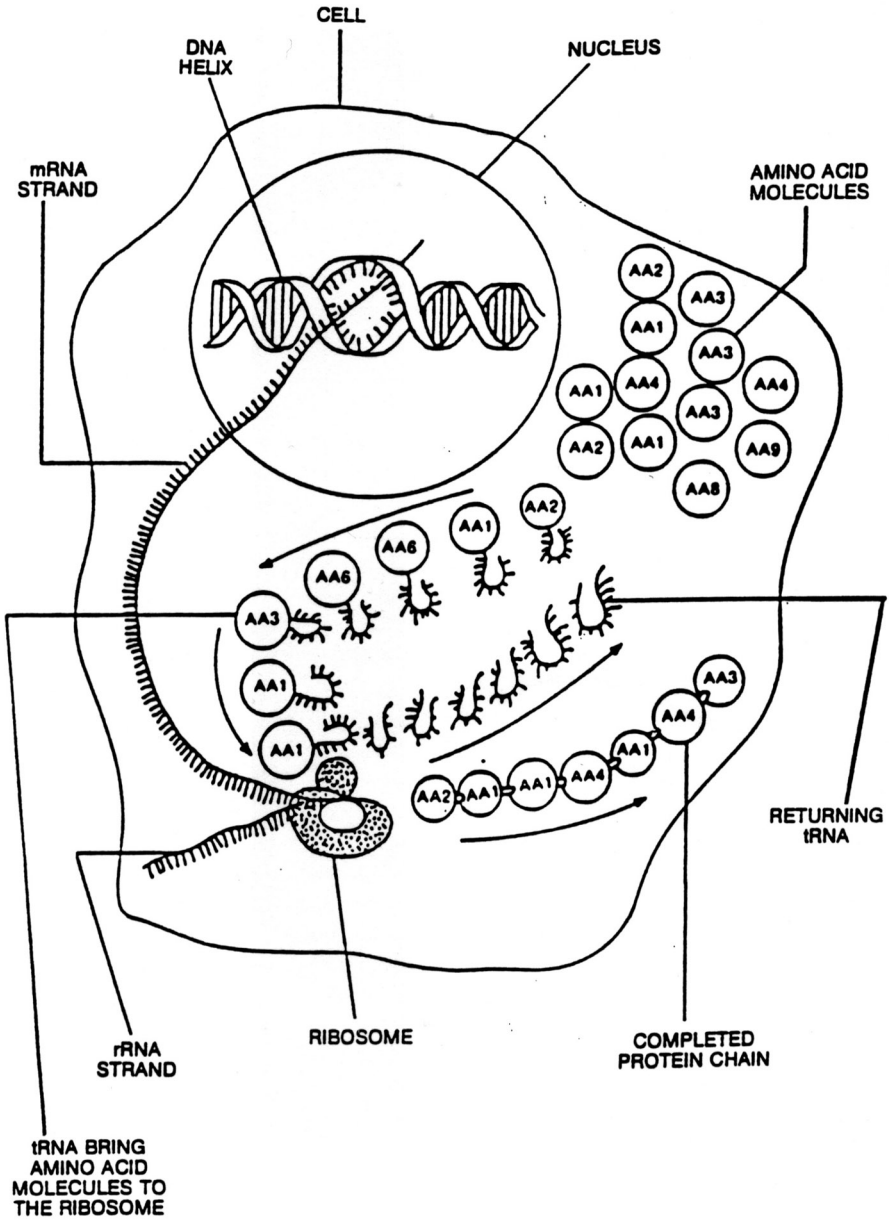

FIGURE 1. The intricate machinery of protein synthesis. The genetic messages encoded on the DNA molecule are copied and then transported by messenger RNA to the ribosome complex. There the genetic message is "read" and translated with the aid of other large biomolecules (transfer-RNA and specific enzyme) to produce a growing amino acid chain. Courtesy of I. L. Cohen of New Research Publications.

During the 1940s, Claude Shannon at Bell Laboratories developed a mathematical theory of information.[27] His theory equated the amount of information transmitted with the amount of uncertainty reduced or eliminated by a series of symbols or characters.[28] For example, before one rolls a six-sided die, there are six possible outcomes. Before one flips a coin, there are two. Rolling a die will thus eliminate more uncertainty and, on Shannon's theory, will convey more information than flipping a coin. Equating information with the reduction of uncertainty implied a mathematical relationship between information and probability (or its inverse, complexity). Note that for a die each possible outcome has only a one in six chance of occurring, compared to a one in two chance for each side of the coin. Thus, in Shannon's theory the occurrence of the more improbable event conveys more information. Shannon generalized this relationship by stating that the amount of information conveyed by an event is inversely proportional to the prior probability of its occurrence. The greater the number of possibilities, the greater the improbability of any one being actualized, and thus more information is transmitted when a particular possibility occurs.

Moreover, information increases as improbabilities multiply. The probability of getting four heads in a row when flipping a fair coin is $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$, or $(\frac{1}{2})^4$. Thus, the probability of attaining a specific sequence of heads and/or tails decreases exponentially as the number of trials increases. The quantity of information increases correspondingly. Even so, information theorists found it convenient to measure information additively rather than multiplicatively. Thus, the common mathematical expression ($I = -\log_2 p$) for calculating information converts probability values into informational measures through a negative logarithmic function, where the negative sign expresses an inverse relationship between information and probability.[29]

Shannon's theory applies most easily to sequences of alphabetic symbols or characters that function as such. Within any given alphabet of x possible characters, the placement of a specific character eliminates x-1 other possibilities and thus a corresponding amount of uncertainty. Or put differently, within any given alphabet or ensemble of x possible characters (where each character has an equi-probable chance of occurring), the probability of any one character occurring is 1/x. The larger the value of x, the greater the amount of information that is conveyed by the occurrence of a specific character in a sequence. In systems where the value of x can be known (or estimated), as in a code or language, mathematicians can easily generate

quantitative estimates of information-carrying capacity. The greater the number of possible characters at each site and the longer the sequence of characters, the greater is the information-carrying capacity—or Shannon information—associated with the sequence.

The essentially digital character of the nucleotide bases in DNA and of the amino acid residues in proteins enabled molecular biologists to calculate the information-carrying capacity (or syntactic information) of those molecules using the new formalism of Shannon's theory. Because at every site in a growing amino acid chain, for example, the chain may receive any one of twenty amino acids, placement of a single amino acid in the chain eliminates a quantifiable amount of uncertainty and increases the Shannon or syntactic information of a polypeptide by a corresponding amount. Similarly, since at any given site along the DNA backbone any one of four nucleotide bases may occur with equal probability, the p value for the occurrence of a specific nucleotide at that site equals 1/4, or .25.[30] The information-carrying capacity of a sequence of a specific length n can then be calculated using Shannon's familiar expression ($I = -\log_2 p$) once one computes a p value for the occurrence of a particular sequence n nucleotides long where $p = (1/4)^n$. The p value thus yields a corresponding measure of information-carrying capacity or syntactic information for a sequence of n nucleotide bases.[31]

E. Complexity, Specificity, and Biological Information

Though Shannon's theory and equations provided a powerful way to measure the amount of information that could be transmitted across a communication channel, it had important limits. In particular, it did not and could not distinguish merely improbable sequences of symbols from those that conveyed a message. As Warren Weaver made clear in 1949, "The word *information* in this theory is used in a special mathematical sense that must not be confused with its ordinary usage. In particular, information must not be confused with meaning."[32] Information theory could measure the information-carrying capacity or the syntactic information of a given sequence of symbols but could not distinguish the presence of a meaningful or functional arrangement of symbols from a random sequence (for example, "we hold these truths to be self-evident" versus "ntnyhiznlhteqkhgdsjh"). Thus, Shannon information theory could quantify the amount of functional or meaningful information that *might be present* in a given sequence of symbols or characters, but it could not distinguish the

status of a functional or message-bearing text from random gibberish. Thus, paradoxically, random sequences of letters often have more syntactic information (or information-carrying capacity), as measured by classical information theory, than do meaningful or functional sequences that happen to contain a certain amount of intentional redundancy or repetition.

In essence, therefore, Shannon's theory remains silent on the important question of whether a sequence of symbols is functionally specific or meaningful. Nevertheless, in its application to molecular biology, Shannon information theory did succeed in rendering rough quantitative measures of the information-carrying capacity or syntactic information (where those terms correspond to measures of brute complexity).[33] As such, information theory did help to refine biologists' understanding of one important feature of the crucial biomolecular components on which life depends: DNA and proteins are highly complex, and quantifiably so. Yet the theory by itself could not establish whether base sequences in DNA or amino acid sequences in proteins possessed the property of functional specificity. Information theory helped establish that DNA and proteins *could* carry large amounts of functional information; it could not establish whether they did.

The ease with which information theory applied to molecular biology (to measure information-carrying capacity) has created considerable confusion about the sense in which DNA and proteins contain "information." Information theory strongly suggested that such molecules possess vast information-carrying capacities or large amounts of syntactic information, as defined by Shannon's theory. When molecular biologists have described DNA as the carrier of hereditary information, however, they have meant much more than the technically limited term *information*. Instead, as Sahotra Sarkar points out, leading molecular biologists defined biological information so as to incorporate the notion of specificity of function (as well as complexity) as early as 1958.[34] Molecular biologists such as Monod and Crick understood biological information—the information stored in DNA and proteins—as something more than mere complexity (or improbability). Their notion of information did associate both biochemical contingency and combinatorial complexity with DNA sequences (allowing DNA's carrying capacity to be calculated), but they also recognized that sequences of nucleotides and amino acids in functioning biomacromolecules possessed a high degree of *specificity* relative to the maintenance of cellular function. As Crick explained in 1958, "By information I mean the specification of the amino acid sequence in protein. . . . Information means here

the *precise* determination of sequence, either of bases in the nucleic acid or on amino acid residues in the protein."[35]

Since the late 1950s, biologists have equated the "*precise* determination of sequence" with the extra-information-theoretic property of specificity or specification. Biologists have defined *specificity* tacitly as "necessary to achieve or maintain function." They have determined that DNA base sequences, for example, are specified not by applying information theory but by making experimental assessments of the function of those sequences within the overall apparatus of gene expression.[36] Similar experimental considerations established the functional specificity of proteins.

Further, developments in complexity theory have now made possible a fully general theoretical account of specification, one that applies readily to biological systems. In particular, recent work by mathematician William Dembski has employed the notion of a rejection region from statistics to provide a formal complexity-theoretic account of specification. According to Dembski, a specification occurs when an event or object (a) falls within an independently given pattern or domain, (b) "matches" or exemplifies a conditionally independent pattern, or (c) meets a conditionally independent set of functional requirements.[37]

To illustrate Dembski's notion of specification, consider these two strings of characters:

"iuinsdysk]idfawqnzkl,mfdifhs"
"Time and tide wait for no man."

Given the number of possible ways of arranging the letters and punctuation marks of the English language for sequences of this length, both of these two sequences constitute highly improbable arrangements of characters. Thus, both have a considerable and quantifiable information-carrying capacity. Nevertheless, only the second of the two sequences exhibits a specification on Dembski's account. To see why, consider the following. Within the set of combinatorially possible sequences, only a very few will convey meaning. This smaller set of meaningful sequences, therefore, delimits a domain or pattern within the larger set of the totality of possibilities. Moreover, this set constitutes a "conditionally independent" pattern. Roughly speaking, a conditionally independent pattern corresponds to a preexisting pattern or set of functional requirements, not one contrived after the fact of observing the event in question—specifically, in this case, the event of observing the two sequences above.[38] Since the smaller domain distinguishes functional from nonfunctional English sequences and

the functionality of alphabetic sequences depends on the preexisting or independently given conventions of English vocabulary and grammar, the smaller set or domain qualifies as a conditionally independent pattern.[39] Since the second string of characters ("Time and tide wait . . .") falls within this smaller conditionally independent domain (or "matches" one of the possible meaningful sentences that fall within it), the second sequence exhibits a specification according to Dembski's complexity-theoretic account. That sequence therefore exhibits the joint properties of complexity and specification and possesses not just information-carrying capacity but both "specified" and, in this case, "semantic" information.

Biological organisms also exhibit specifications, though not necessarily semantic or subjectively "meaningful" ones. The nucleotide base sequences in the coding regions of DNA are highly specific relative to the independent functional requirements of protein function, protein synthesis, and cellular life. To maintain viability, the cell must regulate its metabolism, pass materials back and forth across its membranes, destroy waste materials, and do many other specific tasks. Each of these functional requirements in turn necessitates specific molecular constituents, machines, or systems (usually made of proteins) to accomplish these tasks. Building these proteins with their specific three-dimensional shapes requires specific arrangements of nucleotide bases on the DNA molecule.

Since the chemical properties of DNA allow a vast ensemble of combinatorially possible arrangements of nucleotide bases, any particular sequence will necessarily be highly improbable and rich in Shannon information or information-carrying capacity. Yet within that set of possible sequences a very few will, given the multimolecular system of gene expression within the cell, produce functional proteins.[40] Those that do are thus not only improbable but also functionally "specified" or "specific," as molecular biologists use the terms. Indeed, the smaller set of functionally efficacious sequences again delimits a domain or pattern within a larger set of combinatorial possibilities. Moreover, this smaller domain constitutes a conditionally independent pattern, since (as with the English sequences above) it distinguishes functional from nonfunctional sequences, and the functionality of nucleotide base sequences depends on the independent requirements of protein function. Thus, any actual nucleotide sequence that falls within this domain (or "matches" one of the possible functional sequences that fall within it) exhibits a specification. Put differently, any nucleotide base sequence that produces a functional protein clearly meets certain independent functional requirements, in particular, those of

protein function. Thus, any sequence that meets such requirements (or "falls within the smaller subset of functional sequences") is again not only highly improbable but also specified relative to that independent pattern or domain. Thus, the nucleotide sequences in the coding regions of DNA possess both syntactic information and "specified" information.

A note of definitional clarity must be offered about the relationship between "specified" information and "semantic" information. Though natural languages and DNA base sequences are both specified, only natural language conveys meaning. If one defines "semantic information" as "subjectively meaningful information that is conveyed syntactically (as a string of phonemes or characters) and is understood by a conscious agent," then clearly the information in DNA does not qualify as semantic. Unlike a written or spoken natural language, DNA does not convey "meaning" to a conscious agent.

Rather, the coding regions of DNA function in much the same way as a software program or machine code, directing operations within a complex material system via highly complex yet specified sequences of characters. As Richard Dawkins has noted, "The machine code of the genes is uncannily computer-like."[41] Or as software developer Bill Gates has noted, "DNA is like a computer program, but far, far more advanced than any software we've ever created."[42] Just as the specific arrangement of two symbols (0 and 1) in a software program can perform a function within a machine environment, so, too, can the precise sequencing of the four nucleotide bases in DNA perform a function within the cell.

Though DNA sequences do not convey "meaning," they do exhibit specificity or specification. Moreover, as in a machine code, the sequence specificity of DNA occurs within a syntactic (or functionally alphabetic) domain. Thus, DNA possesses both syntactic and specified information. In any case, since the late 1950s, the concept of information as employed by molecular biologists has conjoined the notions of complexity (or improbability) and specificity of function. The crucial biomolecular constituents of living organisms possess not only Shannon or syntactic information but also "*specified* information" or "*specified* complexity."[43] Biological information so defined, therefore, constitutes a salient feature of living systems that any origin-of-life scenario must explain "the origin of." Further, as we will see below, all naturalistic chemical evolutionary theories have encountered difficulty explaining the origin of such functionally "specified" biological information.

F. Information as Metaphor: Nothing to Explain?

Though most molecular biologists would see nothing controversial in characterizing DNA and proteins as "information-bearing" molecules, some historians and philosophers of biology have recently challenged that description. Before evaluating competing types of explanation for the origin of biological information, this challenge must be addressed. In 2000, the late historian of science Lily Kay characterized the application of information theory to biology as a failure, in particular because classical information theory could not capture the idea of meaning. She suggests, therefore, that the term *information* as used in biology constitutes nothing more than a metaphor. Since, in Kay's view, the term does not designate anything real, it follows that the origin of "biological information" does not require explanation. Instead, only the origin of the *use* of the term *information* within biology requires explanation. As a social constructivist, Kay explained this usage as the result of various social forces operating within the "Cold War Technoculture."[44] In a different but related vein, Sarkar has argued that the concept of information has little theoretical significance in biology because it lacks predictive or explanatory power.[45] He, like Kay, seems to regard the concept of information as a superfluous metaphor lacking empirical reference and ontological status.

Of course, insofar as the term *information* connotes semantic meaning, it does function as a metaphor within biology. That does not mean, however, that the term functions *only* metaphorically or that origin-of-life biologists have nothing to explain. Though information theory has a *limited* application in describing biological systems, it has succeeded in rendering quantitative assessments of the complexity of biomacromolecules. Further, experimental work established the functional specificity of the sequences of monomers in DNA and proteins. Thus, the term *information* as used in biology does refer to two real and contingent properties of living systems: complexity and specificity. Indeed, since scientists began to think seriously about what would be required to explain the phenomenon of heredity, they have recognized the need for some feature or substance in living organisms possessing precisely these two properties together. Thus, Schrödinger envisioned an "aperiodic crystal"; Chargaff perceived DNA's capacity for "complex sequencing"; Watson and Crick equated complex sequences with "information," which Crick in turn equated with "specificity"; Monod equated irregular specificity in proteins with the need for "a code"; and Orgel characterized life as a "specified complexity."[46] Further, Davies has recently argued that the "specific randomness" of DNA base

sequences constitutes the central mystery surrounding the origin of life.[47] Whatever the terminology, scientists have recognized the need for, and now know the location of, a source of complex specificity in the cell in order to transmit heredity and maintain biological function. The incorrigibility of these descriptive concepts suggests that complexity and specificity constitute real properties of biomacromolecules—indeed, properties that could be otherwise but only to the detriment of cellular life. As Orgel notes: "Living organisms are distinguished by their specified complexity. Crystals . . . fail to qualify as living because they lack complexity; mixtures of random polymers fail to qualify because they lack specificity."[48]

The origin of specificity and complexity (in combination), to which the term *information* in biology commonly refers, therefore does require explanation, even if the concept of information connotes only complexity in classical information theory and even if it has no explanatory or predictive value in itself. Instead, as a descriptive (rather than as an explanatory or predictive) concept, the term *information* helps to define (either in conjunction with the notion of "specificity" or by subsuming it) the effect that origin-of-life researchers must explain "the origin of." Thus, *only* where information connotes subjective meaning does it function as a metaphor in biology. Where it refers to an analog of meaning, namely, functional specificity, it defines an essential feature of living systems.

II.

A. Naturalistic Explanations for the Origin of Specified Biological Information

The discoveries of molecular biologists during the 1950s and 1960s raised the question of the ultimate origin of the specified complexity or specified information in both DNA and proteins. Since at least the mid-1960s, many scientists have regarded the origin of information (so defined) as the central question facing origin-of-life biology.[49] Accordingly, origin-of-life researchers have proposed three broad types of naturalistic explanation to explain the origin of specified genetic information: those emphasizing chance, necessity, or the combination of the two.

B. Beyond the Reach of Chance

Perhaps the most common popular naturalistic view about the origin of life is that it happened exclusively by chance. A few serious scientists have also voiced support for this view, at least, at various points in their careers. In

1954, biochemist George Wald, for example, argued for the causal efficacy of chance in conjunction with vast expanses of time. As he explained, "Time is in fact the hero of the plot. . . . Given so much time, the impossible becomes possible, the possible probable, and the probable virtually certain."[50] Later, in 1968, Francis Crick would suggest that the origin of the genetic code—that is, the translation system—might be a "frozen accident."[51] Other theories have invoked chance as an explanation for the origin of genetic information, though often in conjunction with prebiotic natural selection (see part C below).

Almost all serious origin-of-life researchers now consider "chance" an inadequate causal explanation for the origin of biological information.[52] Since molecular biologists began to appreciate the sequence specificity of proteins and nucleic acids in the 1950s and 1960s, many calculations have been made to determine the probability of formulating functional proteins and nucleic acids at random. Various methods of calculating probabilities have been offered by Morowitz, Hoyle and Wickramasinghe, Cairns-Smith, Prigogine, Yockey, and, more recently, Robert Sauer.[53] For the sake of argument, these calculations have often assumed extremely favorable prebiotic conditions (whether realistic or not), much more time than was actually available on the early earth, and theoretically maximal reaction rates among constituent monomers (that is, the constituent parts of proteins, DNA, or RNA). Such calculations have invariably shown that the probability of obtaining functionally sequenced biomacromolecules at random is, in Prigogine's words, "vanishingly small . . . even on the scale of . . . billions of years."[54] As Cairns-Smith wrote in 1971: "Blind chance . . . is very limited. Low-levels of cooperation he [blind chance] can produce exceedingly easily (the equivalent of letters and small words), but he becomes very quickly incompetent as the amount of organization increases. Very soon indeed long waiting periods and massive material resources become irrelevant."[55]

Consider the probabilistic hurdles that must be overcome to construct even one short protein molecule of 100 amino acids in length. (A typical protein consists of about 300 amino acid residues, and many crucial proteins are much longer.)

First, all amino acids must form a chemical bond known as a peptide bond when joining with other amino acids in the protein chain. Yet in nature many other types of chemical bonds are possible between amino acids; in fact, peptide and nonpeptide bonds occur with roughly equal probability. Thus, at any given site along a growing amino acid chain, the probability of

having a peptide bond is roughly ½. The probability of attaining four peptide bonds is (½ × ½ × ½ × ½) = 1/16, or $(½)^4$. The probability of building a chain of 100 amino acids in which all linkages involve peptide linkages is $(½)^{99}$, or roughly 1 chance in 10^{30}.

Second, in nature, every amino acid found in proteins (with one exception) has a distinct mirror image of itself, one left-handed version, or L-form, and one right-handed version, or D-form. These mirror-image forms are called optical isomers. Functioning proteins tolerate only left-handed amino acids, yet the right-handed and left-handed isomers are produced in (amino acid–producing) chemical reactions with roughly equal frequency. Taking this "chirality" into consideration compounds the improbability of attaining a biologically functioning protein. The probability of attaining at random only L-amino acids in a hypothetical peptide chain 100 amino acids long is $(½)^{100}$ or again roughly 1 chance in 10^{30}. Starting from mixtures of DL- forms, the probability of building a 100-amino-acid-length chain at random in which all bonds are peptide bonds and all amino acids are L-form is, therefore, roughly 1 chance in 10^{60}.

Functioning proteins have a third independent requirement, the most important of all; their amino acids must link up in a specific sequential arrangement just as the letters in a meaningful sentence must. In some cases, changing even one amino acid at a given site results in loss of protein function. Moreover, because there are twenty biologically occurring amino acids, the probability of getting a specific amino acid at a given site is small—1/20. (Actually the probability is even lower because in nature there are also many nonprotein-forming amino acids.) On the assumption that all sites in a protein chain require one particular amino acid, the probability of attaining a particular protein 100 amino acids long would be $(1/20)^{100}$, or roughly 1 chance in 10^{130}. We know now, however, that some sites along the chain do tolerate several of the twenty amino acids commonly found in proteins, though others do not. Biochemist Robert Sauer of MIT has used a technique known as "cassette mutagenesis" to determine how much variance among amino acids can be tolerated at any given site in several proteins. His results imply that, even taking the possibility of variance into account, the probability of achieving a functional sequence of amino acids in several known (roughly 100 residue) proteins at random is still "vanishingly small," about 1 chance in 10^{65}.[56] (There are 10^{65} atoms in our galaxy).[57] Recently, Douglas Axe of Cambridge University has used a refined mutagenesis technique to measure the sequence specificity of the protein barnase, a bacterial RNase. Axe's work suggests that previous

mutagenesis experiments actually underestimated the functional sensitivity of proteins to amino acid sequence change because they presupposed (incorrectly) the context independence of individual residue changes.[58] If, in addition to the improbability of attaining a proper sequence, one considers the need for proper bonding and homochirality, the probability of constructing a rather short functional protein at random becomes so small (no more than 1 chance in 10^{125}) as to appear absurd on the chance hypothesis. As Dawkins has said, "We can accept a certain amount of luck in our explanations, but not too much."[59]

Of course, Dawkins's assertion begs a quantitative question, namely, "How improbable does an event, sequence, or system have to be before the chance hypothesis can be reasonably eliminated?" That question has recently received a formal answer. William Dembski, following and refining the work of earlier probabilists such as Emile Borel, has shown that chance can be eliminated as a plausible explanation for specified systems of small probability whenever the complexity of a specified event or sequence exceeds available probabilistic resources.[60] He then calculates a conservative estimate for the "universal probability bound" of 1 in 10^{150} corresponding to the probabilistic resources of the known universe. This number provides a theoretical basis for excluding appeals to chance as the best explanation for specified events of probability less than $\frac{1}{2} \times 10^{150}$. Dembski thus answers the question of how much luck is—for any case—too much to invoke in an explanation.

Significantly, the improbability of assembling and sequencing even a short functional protein approaches this universal probability bound—the point at which appeals to chance become absurd given the "probabilistic resources" of the entire universe.[61] Further, making the same kind of calculation for even moderately longer proteins pushes these measures of improbability well beyond the limit. For example, the probability of generating a protein of only 150 amino acids in length (using the same method as above) is less than 1 chance in 10^{180}, well beyond the most conservative estimates of the probability bound, given our multibillion year old universe.[62] Thus, given the complexity of proteins, it is extremely unlikely that a random search through the space of combinatorially possible amino acid sequences could generate even a single relatively short functional protein in the time available since the beginning of the universe (let alone the time available on the early earth). Conversely, to have a reasonable chance of finding a short functional protein in a random search of

combinatorial space would require vastly more time than either cosmology or geology allows.

More realistic calculations (taking into account the probable presence of nonproteineous amino acids, the need for much longer proteins to perform specific functions such as polymerization, and the need for hundreds of proteins working in coordination to produce a functioning cell) only compound these improbabilities, almost beyond computability. For example, recent theoretical and experimental work on the so-called minimal complexity required to sustain the simplest possible living organism suggests a lower bound of some 250 to 400 genes and their corresponding proteins.[63] The nucleotide sequence-space corresponding to such a system of proteins exceeds $4^{300,000}$. The improbability corresponding to this measure of molecular complexity again vastly exceeds 1 chance in 10^{150} and thus the "probabilistic resources" of the entire universe.[64] When one considers the full complement of functional biomolecules required to maintain minimal cell function and vitality, one can see why chance-based theories of the origin of life have been abandoned. What Mora said in 1963 still holds: "Statistical considerations, probability, complexity, etc., followed to their logical implications suggest that the origin and continuance of life is not controlled by such principles. An admission of this is the use of a period of practically infinite time to obtain the derived result. Using such logic, however, we can prove anything."[65]

Though the probability of assembling a functioning biomolecule or cell by chance alone is exceedingly small, it is important to emphasize that scientists have not generally rejected the chance hypothesis merely because of the vast improbabilities associated with such events. Very improbable things do occur by chance. Any hand of cards or any series of rolled dice will represent a highly improbable occurrence. Observers often justifiably attribute such events to chance alone. What justifies the elimination of chance is not just the occurrence of a highly improbable event but also the occurrence of an improbable event that also conforms to a discernible pattern (that is, to a conditionally independent pattern; see part I, section E). If someone repeatedly rolls two dice and turns up a sequence such as 9, 4, 11, 2, 6, 8, 5, 12, 9, 2, 6, 8, 9, 3, 7, 10, 11, 4, 8, and 4, no one will suspect anything but the interplay of random forces, though this sequence does represent a very improbable event given the number of combinatorial possibilities that correspond to a sequence of this length. Yet rolling 20 (or certainly 200) consecutive sevens will justifiably arouse suspicion that something more than chance is in play. Statisticians have long used a

method for determining when to eliminate the chance hypothesis; the method requires prespecifying a pattern or "rejection region."[66] In the dice example above, one could prespecify the repeated occurrence of seven as such a pattern in order to detect the use of loaded dice, for example. Dembski has generalized this method to show how the presence of any conditionally independent pattern, whether temporally prior to the observation of an event or not, can help (in conjunction with a small probability event) to justify rejecting the chance hypothesis.[67]

Origin-of-life researchers have tacitly, and sometimes explicitly, employed this kind of statistical reasoning to justify the elimination of scenarios relying heavily on chance. Christian de Duve, for example, has made the logic explicit in order to explain why chance fails as an explanation for the origin of life: "A single, freak, highly improbable event can conceivably happen. Many highly improbable events—drawing a winning lottery number or the distribution of playing cards in a hand of bridge—happen all the time. But a string of improbable events—drawing the same lottery number twice, or the same bridge hand twice in a row—does not happen naturally."[68]

De Duve and other origin-of-life researchers have long recognized that the cell represents not only a highly improbable but also a functionally specified system. For this reason, by the mid-1960s most researchers had eliminated chance as a plausible explanation for the origin of the specified information necessary to build a cell.[69] Many have instead sought other types of naturalistic explanations.

C. Prebiotic Natural Selection: A Contradiction in Terms

Of course, even many early theories of chemical evolution did not rely *exclusively* on chance as a causal mechanism. For example, Oparin's original theory of evolutionary abiogenesis first published in the 1920s and 1930s invoked prebiotic natural selection as a complement to chance interactions. Oparin's theory envisioned a series of chemical reactions that he thought would enable a complex cell to assemble itself gradually and naturalistically from simple chemical precursors.

For the first stage of chemical evolution, Oparin proposed that simple gases such as ammonia (NH_3), methane (CH_4), water vapor (H_2O), carbon dioxide (CO_2), and hydrogen (H_2)would have existed in contact with the early oceans and with metallic compounds extruded from the core of the earth.[70] With the aid of ultraviolet radiation from the sun, the ensuing

reactions would have produced energy-rich hydrocarbon compounds. They in turn would have combined and recombined with various other compounds to make amino acids, sugars, and other "building blocks" of complex molecules such as proteins necessary to living cells. These constituents would eventually arrange themselves by chance into primitive metabolic systems within simple cell-like enclosures that Oparin called coacervates. Oparin then proposed a kind of Darwinian competition for survival among his coacervates. Those that, by chance, developed increasingly complex molecules and metabolic processes would have survived to grow more complex and efficient. Those that did not would have dissolved.[71] Thus, Oparin invoked differential survival or natural selection as a mechanism for preserving complexity-increasing events, thus allegedly helping to overcome the difficulties attendant to pure-chance hypotheses.

Developments in molecular biology during the 1950s cast doubt on Oparin's scenario. Oparin originally invoked natural selection to explain how cells refined primitive metabolism once it had arisen. His scenario relied heavily on chance to explain the initial formation of the constituent biomacromolecules on which even primitive cellular metabolism would depend. Discovery during the 1950s of the extreme complexity and specificity of such molecules undermined the plausibility of his claim. For that and other reasons, Oparin published a revised version of his theory in 1968 that envisioned a role for natural selection earlier in the process of abiogenesis. His new theory claimed that natural selection acted on random polymers as they formed and changed within his coacervate protocells.[72] As more complex and efficient molecules accumulated, they would have survived and reproduced more prolifically.

Even so, Oparin's concept of *prebiotic* natural selection acting on initially unspecified biomacromolecules remained problematic. For one thing, it seemed to presuppose a preexisting mechanism of self-replication. Yet self-replication in all extant cells depends on functional and, therefore, (to a high degree) sequence-specific proteins and nucleic acids. Yet the origin of specificity in these molecules is precisely what Oparin needed to explain. As Christian de Duve has stated, theories of prebiotic natural selection "need information which implies they have to presuppose what is to be explained in the first place."[73] Oparin attempted to circumvent the problem by claiming that the first polymers need not have been highly sequence-specific. But that claim raised doubts about whether an accurate mechanism of self-replication (and thus natural selection) could have functioned at all. Oparin's latter scenario did not reckon on a phenomenon

known as error catastrophe, in which small errors, or deviations from functionally necessary sequences, are quickly amplified in successive replications.[74]

Thus, the need to explain the origin of specified information created an intractable dilemma for Oparin. On the one hand, if he invoked natural selection late in his scenario, he would need to rely on chance alone to produce the highly complex and specified biomolecules necessary to self-replication. On the other hand, if Oparin invoked natural selection earlier in the process of chemical evolution, before functional specificity in biomacromolecules would have arisen, he could give no account of how such prebiotic natural selection could even function (given the phenomenon of error-catastrophe). Natural selection presupposes a self-replication system, but self-replication requires functioning nucleic acids and proteins (or molecules approaching their complexity)—the very entities that Oparin needed to explain. Thus, Dobzhansky would insist that, "prebiological natural selection is a contradiction in terms."[75]

Although some rejected the hypothesis of prebiotic natural selection as question-begging, others dismissed it as indistinguishable from implausible chance-based hypotheses.[76] The work of mathematician John von Neumann supported that judgment. During the 1960s, von Neumann showed that any system capable of self-replication would require subsystems that were functionally equivalent to the information storage, replicating, and processing systems found in extant cells.[77] His calculations established a very high minimal threshold of biological function, as would later experimental work.[78] These minimal-complexity requirements pose a fundamental difficulty for natural selection. Natural selection selects for functional advantage. It can play no role, therefore, until random variations produce some biologically advantageous arrangement of matter. Yet von Neumann's calculations and similar ones by Wigner, Landsberg, and Morowitz showed that in all probability (to understate the case) random fluctuations of molecules would not produce the minimal complexity needed for even a primitive replication system.[79] As noted above, the improbability of developing a functionally integrated replication system vastly exceeds the improbability of developing the protein or DNA components of such a system. Given the huge improbability and the high functional threshold it implies, many origin-of-life researchers came to regard prebiotic natural selection as both inadequate and essentially indistinguishable from appeals to chance.

Nevertheless, during the 1980s, Richard Dawkins and Bernd-Olaf Kuppers attempted to resuscitate prebiotic natural selection as an explanation for the origin of biological information.[80] Both accept the futility of naked appeals to chance and invoke what Kuppers calls a "Darwinian optimization principle." Both use computers to demonstrate the efficacy of prebiotic natural selection. Each selects a target sequence to represent a desired functional polymer. After creating a crop of randomly constructed sequences and generating variations among them at random, their computers select those sequences that match the target sequence most closely. The computers then amplify the production of those sequences, eliminate the others (to simulate differential reproduction), and repeat the process. As Kuppers puts it, "Every mutant sequence that agrees one bit better with the meaningful or reference sequence . . . will be allowed to reproduce more rapidly."[81] In his case, after a mere thirty-five generations, his computer succeeded in spelling his target sequence, "NATURAL SELECTION."

Despite superficially impressive results, such "simulations" conceal an obvious flaw: Molecules in situ do not have a target sequence "in mind." Nor will they confer any selective advantage on a cell, and thus differentially reproduce, until they combine in a functionally advantageous arrangement. Thus, nothing in nature corresponds to the role that the computer plays in selecting functionally nonadvantageous sequences that happen to agree "one bit better" than others with a target sequence. The sequence NORMAL ELECTION may agree more with NATURAL SELECTION than does the sequence MISTRESS DEFECTION, but neither of the two yields any advantage in communication over the other in trying to communicate something about NATURAL SELECTION. If that is the goal, both are equally ineffectual. Even more to the point, a completely nonfunctional polypeptide would confer no selective advantage on a hypothetical protocell, even if its sequence happened to agree "one bit better" with an unrealized target protein than some other nonfunctional polypeptide.

Both Kuppers's and Dawkins's published results of their simulations show the early generations of variant phrases awash in nonfunctional gibberish.[82] In Dawkins's simulation, not a single functional English word appears until after the tenth iteration (unlike the more generous example above that starts with actual, albeit incorrect, words). Yet to make distinctions on the basis of function among sequences that have no function is entirely unrealistic. Such determinations can be made only if considerations of *proximity to possible future function* are allowed, but that requires foresight, which natural selection does not have. A computer, programmed

by a human being, can perform such functions. To imply that molecules can do so as well illicitly personifies nature. Thus, if these computer simulations demonstrate anything, they subtly demonstrate the need for intelligent agents to elect some options and exclude others—that is, to create information.

D. Self-Organizational Scenarios

Because of the difficulties with chance-based theories, including those relying on prebiotic natural selection, most origin-of-life theorists after the mid-1960s attempted to address the problem of the origin of biological information in a completely different way. Researchers began to look for self-organizational laws and properties of chemical attraction that might explain the origin of the specified information in DNA and proteins. Rather than invoking chance, such theories invoked necessity. If neither chance nor prebiotic natural selection acting on chance explains the origin of specified biological information, then those committed to finding a naturalistic explanation for the origin of life must necessarily rely on physical or chemical necessity. Given a limited number of broad explanatory categories, the inadequacy of chance (with or without prebiotic natural selection) has, in the minds of many researchers, left only one option. Christian de Duve articulates the logic: "a string of improbable events—drawing the same lottery number twice, or the same bridge hand twice in a row—does not happen naturally. All of which lead me to conclude that life is an obligatory manifestation of matter, bound to arise where conditions are appropriate."[83]

When origin-of-life biologists began considering the self-organizational perspective that de Duve describes, several researchers proposed that deterministic forces (stereochemical "necessity") made the origin of life not just probable but inevitable. Some suggested that simple chemicals possessed "self-ordering properties" capable of organizing the constituent parts of proteins, DNA, and RNA into the specific arrangements they now possess.[84] Steinman and Cole, for example, suggested that differential bonding affinities or forces of chemical attraction between certain amino acids might account for the origin of the sequence specificity of proteins.[85] Just as electrostatic forces draw sodium (Na+) and chloride (Cl-) ions together into highly ordered patterns within a crystal of salt (NaCl), so, too, might amino acids with special affinities for each other arrange themselves to form proteins. In 1969, Kenyon and Steinman developed that idea in a book entitled *Biochemical Predestination.* They argued that life might have been "biochemically predestined" by the properties of attraction existing

between its constituent chemical parts, particularly among the amino acids in proteins.[86]

In 1977, another self-organizational theory was proposed by Prigogine and Nicolis based on a thermodynamic characterization of living organisms. In *Self Organization in Nonequilibrium Systems,* Prigogine and Nicolis classified living organisms as open, nonequilibrium systems capable of "dissipating" large quantities of energy and matter into the environment.[87] They observed that open systems driven far from equilibrium often display self-ordering tendencies. For example, gravitational energy will produce highly ordered vortices in a draining bathtub; thermal energy flowing through a heat sink will generate distinctive convection currents or "spiral wave activity." Prigogine and Nicolis argued that the organized structures observed in living systems might have similarly "self-originated" with the aid of an energy source. In essence, they conceded the improbability of simple building blocks arranging themselves into highly ordered structures under normal equilibrium conditions. But they suggested that, under nonequilibrium conditions, where an external source of energy is supplied, biochemical building blocks might arrange themselves into highly ordered patterns.

More recently, Kauffman and de Duve have proposed self-organizational theories with somewhat less specificity, at least with regard to the problem of the origin of specified genetic information.[88] Kauffman invokes so-called autocatalytic properties to generate metabolism directly from simple molecules. He envisions such autocatalysis occurring once very particular configurations of molecules have arisen in a rich "chemical minestrone." De Duve also envisions protometabolism emerging first with genetic information arising later as a byproduct of simple metabolic activity.

E. Order versus Information

For many current origin-of-life scientists, self-organizational models now seem to offer the most promising approach to explaining the origin of specified biological information. Nevertheless, critics have called into question both the plausibility and the relevance of self-organizational models. Ironically, a prominent early advocate of self-organization, Dean Kenyon, has now explicitly repudiated such theories as both incompatible with empirical findings and theoretically incoherent.[89]

First, empirical studies have shown that some differential affinities do exist between various amino acids (that is, certain amino acids do form linkages more readily with some amino acids than with others).[90] Nevertheless,

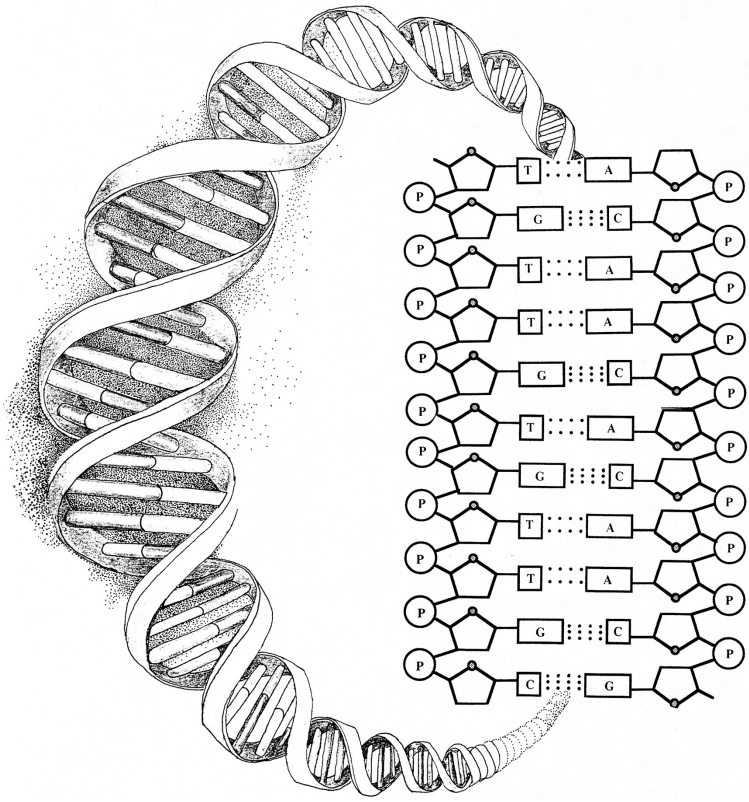

FIGURE 2. The bonding relationship between the chemical constituents of the DNA molecule. Sugars (designated by the pentagons) and phosphates (designated by the circled Ps) are linked chemically. Nucleotide bases (A's, T's, G's and C's) are bonded to the sugar-phosphate backbones. Nucleotide bases are linked by hydrogen bonds (designated by dotted double or triple lines) across the double helix. But no chemical bonds exist between the nucleotide bases along the message-bearing spine of the helix. Courtesy of Fred Heeren, Day Star publications.

such differences do not correlate to actual sequences in large classes of known proteins.[91] In short, differing chemical affinities do not explain the multiplicity of amino acid sequences existing in naturally occurring proteins or the sequential arrangement of amino acids in any particular protein.

In the case of DNA, this point can be made more dramatically. Figure 2 shows that the structure of DNA depends on several chemical bonds. There are bonds, for example, between the sugar and the phosphate molecules

forming the two twisting backbones of the DNA molecule. There are bonds fixing individual (nucleotide) bases to the sugar-phosphate backbones on each side of the molecule. There are also hydrogen bonds stretching horizontally across the molecule between nucleotide bases, making so-called complementary pairs. The individually weak hydrogen bonds, which in concert hold two complementary copies of the DNA message text together, make replication of the genetic instructions possible. It is important to note, however, that there are *no* chemical bonds between the bases along the longitudinal axis in the center of the helix. Yet it is precisely along this axis of the DNA molecule that the genetic information is stored.

Further, just as magnetic letters can be combined and recombined in any way to form various sequences on a metal surface, so, too, can each of the four bases—A, T, G, and C—attach to any site on the DNA backbone with equal facility, making all sequences equally probable (or improbable). Indeed, there are no significant differential affinities between any of the four bases and the binding sites along the sugar-phosphate backbone. The same type of N-glycosidic bond occurs between the base and the backbone regardless of which base attaches. All four bases are acceptable; none is chemically favored. As Kuppers has noted, "The properties of nucleic acids indicate that all the combinatorially possible nucleotide patterns of a DNA are, from a chemical point of view, equivalent."[92] Thus, "self-organizing" bonding affinities cannot explain the sequentially specific arrangement of nucleotide bases in DNA because (1) there are *no* bonds between bases along the information-bearing axis of the molecule, and (2) there are no *differential* affinities between the backbone and the specific bases that could account for variations in sequence. Because the same holds for RNA molecules, researchers who speculate that life began in an RNA world have also failed to solve the sequence specificity problem—that is, the problem of explaining how information in functioning RNA molecules could have arisen in the first place.

For those who want to explain the origin of life as the result of self-organizing properties intrinsic to the material constituents of living systems, these rather elementary facts of molecular biology have decisive implications. The most obvious place to look for self-organizing properties to explain the origin of genetic information is in the constituent parts of the molecules that carry that information. But biochemistry and molecular biology make clear that forces of attraction between the constituents in DNA, RNA, and proteins do not explain the sequence specificity of these large, information-bearing biomolecules.

The properties of the monomers constituting nucleic acids and proteins simply do not make a particular gene, let alone life as we know it, inevitable. (We know this, in addition to the reasons already stated, because of the many variant polypeptides and gene sequences that exist in nature and that have been synthesized in the laboratory.) Yet if self-organizational scenarios for the origin of biological information are to have any theoretical import, they must claim just the opposite. And that claim is often made, albeit without much specificity. As de Duve has put it, "the processes that generated life" were "highly deterministic," making life as we know it "inevitable" given "the conditions that existed on the prebiotic earth."[93] Yet imagine the most favorable prebiotic conditions. Imagine a pool of all four DNA bases and all necessary sugars and phosphates; would any particular genetic sequence inevitably arise? Given all necessary monomers, would any particular functional protein or gene, let alone a specific genetic code, replication system, or signal transduction circuitry, inevitably arise? Clearly not.

In the parlance of origin-of-life research, monomers are "building blocks," and building blocks can be arranged and rearranged in innumerable ways. The properties of stone blocks do not determine their own arrangement in the construction of buildings. Similarly, the properties of *biological* building blocks do not determine the arrangement of functional polymers. Instead, the chemical properties of the monomers allow a vast ensemble of possible configurations, the overwhelming majority of which have no biological function whatsoever. Functional genes or proteins are no more inevitable, given the properties of their "building blocks," than, for example, the Palace of Versailles was inevitable, given the properties of the stone blocks that were used to construct it. To anthropomorphize, neither bricks and stone, nor letters in a written text, nor nucleotide bases "care" how they are arranged. In each case, the properties of the constituents remain largely indifferent to the many specific configurations or sequences they may adopt, nor do they make any specific structures "inevitable" as self-organizationalists must claim.

Significantly, information theory makes clear that there is a good reason for this. If chemical affinities between the constituents in the DNA determined the arrangement of the bases, such affinities would dramatically diminish the capacity of DNA to carry information. Recall that classical information theory equates the reduction of uncertainty with the transmission of information, whether specified or unspecified. The transmission of information, therefore, requires physical-chemical contingency. As Robert

Stalnaker has noted, "[information] content requires contingency."[94] If, therefore, forces of chemical necessity completely determine the arrangement of constituents in a system, that arrangement will not exhibit complexity or convey information.

Consider, for example, what would happen if the individual nucleotide bases (A, C, G, and T) in the DNA molecule *did* interact by *chemical* necessity (along the information-bearing axis of DNA). Suppose that every time adenine (A) occurred in a growing genetic sequence, it attracted cytosine (C) to it.[95] Suppose every time guanine (G) appeared, thymine (T) followed. If this were the case, the longitudinal axis of DNA would be peppered with repetitive sequences in which A followed C and T followed by G. Rather than a genetic molecule capable of virtually unlimited novelty and characterized by unpredictable and aperiodic sequences, DNA would contain sequences awash in repetition or redundancy—much like the arrangement of atoms in crystals. In a crystal, the forces of mutual chemical attraction do determine, to a very considerable extent, the sequential arrangement of its constituent parts. Hence, sequencing in crystals is highly ordered and repetitive but neither complex nor informative. In DNA, however, where any nucleotide can follow any other, a vast array of novel sequences is possible, corresponding to a multiplicity of possible amino acid sequences and protein functions.

The forces of chemical necessity produce redundancy (roughly, law- or rule-generated repetition) or monotonous order but reduce the capacity to convey information and express novelty. Thus, as chemist Michael Polanyi noted:

> Suppose that the actual structure of a DNA molecule were due to the fact that the bindings of its bases were much stronger than the bindings would be for any other distribution of bases, then such a DNA molecule would have no information content. Its code-like character would be effaced by an overwhelming redundancy. . . . Whatever may be the origin of a DNA configuration, it can function as a code only if its order is not due to the forces of potential energy. It *must be* as physically indeterminate as the sequence of words is on a printed page [emphasis added].[96]

In other words, if chemists had found that bonding affinities between the nucleotides in DNA produced nucleotide sequencing, they also would have found that they had been mistaken about DNA's information-bearing properties. Or, to put the point quantitatively, to the extent that forces of attraction between constituents in a sequence determine the arrangement

of the sequence, to that extent will the information-carrying capacity of the system be diminished or effaced by redundancy.[97] As Dretske has explained: "As p(si) [the probability of a condition or state of affairs] approaches 1, the amount of information associated with the occurrence of si goes to 0. In the limiting case when the probability of a condition or state of affairs is unity [p(si) = 1], no information is associated with, or generated by, the occurrence of si. This is merely another way to say that no information is generated by the occurrence of events for which there are no possible alternatives."[98]

Bonding affinities, to the extent they exist, inhibit the maximization of information because they determine that specific outcomes will follow specific conditions with high probability.[99] Yet information-carrying capacity is maximized when just the opposite situation obtains, namely, when antecedent conditions allow many improbable outcomes.

Of course, as noted in part I, section D, the base sequences in DNA do more than possess information-carrying capacity (or syntactic information) as measured by classical Shannon information theory. These sequences store functionally specified information—that is, they are specified as well as complex. Clearly, however, a sequence cannot be both specified and complex if it is not at least complex. Therefore, self-organizational forces of chemical necessity, which produce redundant order and *preclude* complexity, also preclude the generation of specified complexity (or specified information) as well. Chemical affinities do not generate complex sequences. Thus, they cannot be invoked to explain the origin of information, whether specified or otherwise.

A tendency to conflate the qualitative distinctions between "order" and "complexity" has characterized self-organizational scenarios—whether those that invoke internal properties of chemical attraction or an external organizing force or source of energy. That tendency calls into question the relevance of these scenarios of the origin of life. As Yockey has argued, the accumulation of structural or chemical order does not explain the origin of biological complexity or genetic information. He concedes that energy flowing through a system may produce highly ordered patterns. Strong winds form swirling tornados and the "eyes" of hurricanes; Prigogine's thermal baths do develop interesting convection currents; and chemical elements do coalesce to form crystals. Self-organizational theorists explain well what does not need explaining. What needs explaining in biology is not the origin of order (defined as symmetry or repetition) but the specified information—the highly complex, aperiodic, and specified sequences

that make biological function possible. As Yockey warns: "Attempts to relate the idea of order . . . with biological organization or specificity must be regarded as a play on words that cannot stand careful scrutiny. Informational macromolecules can code genetic messages and therefore can carry information because the sequence of bases or residues is affected very little, if at all, by [self-organizing] physicochemical factors."[100]

In the face of these difficulties, some self-organizational theorists have claimed that we must await the discovery of new natural laws to explain the origin of biological information. As Manfred Eigen has argued, "our task is to find an algorithm, a natural law, that leads to the origin of information."[101] Such a suggestion betrays confusion on two counts. First, scientific laws don't generally produce or cause natural phenomena, they describe them. For example, Newton's law of gravitation described, but did not cause or explain, the attraction between planetary bodies. Second, laws necessarily describe highly deterministic or predictable relationships between antecedent conditions and consequent events. Laws describe highly repetitive patterns in which the probability of each successive event (given the previous event) approaches unity. Yet information sequences are complex, not repetitive—information mounts as *improbabilities* multiply. Thus, to say that scientific laws can produce information is essentially a contradiction in terms. Instead, scientific laws describe (almost by definition) highly predictable and regular phenomena—that is, redundant order, not complexity (whether specified or otherwise).

Though the patterns that natural laws describe display a high degree of regularity, and thus lack the complexity that characterizes information-rich systems, one could argue that we might someday discover a very particular configuration of *initial conditions* that routinely generates high informational states. Thus, while we cannot hope to find a law that describes an information-rich *relationship* between antecedent and consequent variables, we might find a law that describes how a very particular set of initial conditions routinely generates a high information state. Yet even the statement of this hypothetical seems itself to beg the question of the ultimate origin of information, since "a very particular set of initial conditions" sounds precisely like an information-rich—a highly complex and specified—state. In any case, everything we know experientially suggests that the amount of specified information present in a set of antecedent conditions necessarily equals or exceeds that of any system produced from those conditions.

F. Other Scenarios and the Displacement of the Information Problem

In addition to the general categories of explanation already examined, origin-of-life researchers have proposed many more specific scenarios, each emphasizing random variations (chance), self-organizational laws (necessity), or both. Some of those scenarios purport to address the information problem; others attempt to bypass it altogether. Yet on closer examination, even scenarios that appear to alleviate the problem of the origin of specified biological information merely shift the problem elsewhere. Genetic algorithms can "solve" the information problem, but only if programmers provide informative target sequences and selection criteria. Simulation experiments can produce biologically relevant precursors and sequences, but only if experimentalists manipulate initial conditions or select and guide outcomes—that is, only if they add information themselves. Origin-of-life theories can leapfrog the problem altogether, but only by presupposing the presence of information in some other preexisting form.

Any number of theoretical models for the origin of life have fallen prey to this difficulty. For example, in 1964, Henry Quastler, an early pioneer in the application of information theory to molecular biology, proposed a DNA-first model for the origin of life. He envisioned the initial emergence of a system of unspecified polynucleotides capable of primitive self-replication via the mechanisms of complementary base-pairing. The polymers in his system would have, on Quastler's account, initially lacked specificity (which he equated with information).[102] Only later, when his system of polynucleotides had come into association with a fully functional set of proteins and ribosomes, would the specific nucleotide sequences in the polymers take on any functional significance. He likened that process to the random selection of a combination for a lock in which the combination would only later acquire functional significance once particular tumblers had been set to allow the combination to open the lock. In both the biological and the mechanical case, the surrounding context would confer functional specificity on an initially unspecified sequence. Thus, Quastler characterized the origin of information in polynucleotides as an "accidental choice remembered."

Although Quastler's way of conceiving of the origin of specified biological information did allow "a chain of nucleotides [to] become a [functional] system of genes without necessarily suffering any change in structure," it did have an overriding difficulty. It did not account for the origin of the

complexity and specificity of the system of molecules whose association with the initial sequence gave the initial sequence functional significance. In Quastler's combination-lock example, conscious agents chose the tumbler settings that made the initial combination functionally significant. Yet Quastler expressly precluded conscious design as a possibility for explaining the origin of life.[103] Instead, he seemed to suggest that the origin of the biological context—that is, the complete set of functionally specific proteins (and the translation system) necessary to create a "symbiotic association" between polynucleotides and proteins—would arise by chance. He even offered some rough calculations to show that the origin of such a multimolecular context, though improbable, would have been probable enough to expect it to occur by chance in the prebiotic soup. Quastler's calculations now seem extremely implausible in light of the discussion of minimal complexity in part II, section B.[104] More significantly, Quastler "solved" the problem of the origin of complex specificity in nucleic acids only by transferring the problem to an equally complex and specified system of proteins and ribosomes. Whereas, admittedly, *any* polynucleotide sequence would suffice initially, the subsequent proteins and ribosomal material constituting the translation system would have to possess an extreme specificity *relative to the initial polynucleotide sequence* and relative to any protocellular functional requirements. Thus, Quastler's attempt to bypass the sequence specificity problem merely shifted it elsewhere.

Self-organizational models have encountered similar difficulties. For example, chemist J. C. Walton has argued (echoing earlier articles by Mora) that the self-organizational patterns produced in Prigogine-style convection currents do not exceed the organization or structural information represented by the experimental apparatus used to create the currents.[105] Similarly, Maynard-Smith, Dyson, and Shapiro have shown that Eigen's so-called hypercycle model for generating biological information actually shows how information tends to degrade over time.[106] Eigen's hypercycles presuppose a large initial contribution of information in the form of a long RNA molecule and some forty specific proteins and thus do not attempt to explain the ultimate origin of biological information. Moreover, because hypercycles lack an error-free mechanism of self-replication, the proposed mechanism succumbs to various "error-catastrophes" that ultimately diminish, not increase, the (specified) information content of the system over time.

Stuart Kauffman's self-organizational theory also subtly transfers the problem of the origin of information. In *The Origins of Order,* Kauffman

attempts to leapfrog the sequence-specificity problem by proposing a means by which a self-reproducing and metabolic system might emerge directly from a set of "low specificity" catalytic peptides and RNA molecules in a prebiotic soup or "chemical minestrone." Kauffman envisions, as Iris Frey puts it, "a set of catalytic polymers in which no single molecule reproduces itself, but the system as a whole does."[107] Kauffman argues that once a sufficiently diverse set of catalytic molecules had assembled (in which the different peptides performed enough different catalytic functions) the ensemble of individual molecules would spontaneously undergo a kind of phase transition resulting in a self-reproducing metabolic system. Thus, Kauffman argues that metabolism can arise directly without genetic information encoded in DNA.[108]

Nevertheless, Kauffman's scenario does not solve, or bypass, the problem of the origin of biological information. Instead, it either presupposes the existence of unexplained sequence-specificity or it transfers such needed specificity out of view. Kauffman claims that an ensemble of relatively short and low specificity catalytic peptides and RNA molecules would suffice jointly to establish a metabolic system. He defends the biochemical plausibility of his scenario on the grounds that some proteins can perform enzymic functions with low specificity and complexity. He cites proteases such as trypsin that cleave peptide bonds at single amino acid sites and proteins in the clotting cascade that "cleave essentially single target polypeptides" to support his claim.[109]

Yet Kauffman's argument has two problems. First, it does not follow, nor is it the case biochemically, that just because *some* enzymes might function with low specificity, that *all* the catalytic peptides (or enzymes) needed to establish a self-reproducing metabolic cycle could function with similarly low levels of specificity and complexity. Instead, modern biochemistry shows that at least some, and probably many, of the molecules in a closed interdependent system of the type that Kauffman envisions would require high complexity and specificity proteins. Enzymatic catalysis (which his scenario would surely necessitate) invariably requires molecules long enough (at least 50-mers) to form tertiary structures (whether in polynucleotides or polypeptides). Further, these long polymers invariably require very specific three-dimensional geometries (which can in turn derive from sequence-specific arrangements of monomers) in order to catalyze necessary reactions. How do these molecules acquire their specificity of sequencing? Kauffman does not address this question because his illustration incorrectly suggests that he need not do so.

Secondly, it turns out that even the allegedly low specificity molecules that Kauffman cites to illustrate the plausibility of his scenario do not themselves manifest low complexity and specificity. Instead, Kauffman has confused the specificity and complexity of the parts of the polypeptides upon which the proteases act with the specificity and complexity of the proteins (the proteases) that do the enzymatic acting. Though trypsin, for example, acts upon (cleaves) peptide bonds at a relatively simple target (the carboxyl end of two separate amino acids, argenine, and lysine), trypsin itself is a highly complex and specifically-sequenced molecule. Indeed, trypsin is a non-repeating 200+ residue protein that possesses significant sequence-specificity as a condition of its function.[110] Further, it has to manifest significant three-dimensional (geometric) specificity to recognize the specific amino acids argenine and lysine—sites at which it cleaves peptide bonds. By equivocating in his discussion of specificity, Kauffman obscures from view the considerable specificity and complexity requirement of even the proteases he cites to justify his claim that low specificity catalytic peptides will suffice to establish a metabolic cycle. Thus, Kauffman's own illustration properly understood (that is, without equivocating about the relevant locus of specificity), shows that for his scenario to have biochemical plausibility it must *presuppose* the existence of many high complexity and specificity polypeptides and polynucleotides. Where does this information in these molecules come from? Kauffman, again, does not say.

Further, Kauffman must acknowledge (as he seems to in places),[111] that for autocatalysis (for which there is as yet no experimental evidence) to occur, the molecules in the "chemical minestrone" must be held in a very specific spatial-temporal relationship to one another. In other words, for the direct autocatalysis of integrated metabolic complexity to occur, a system of catalytic peptide molecules must first achieve a very specific molecular configuration, or a low configurational entropy state.[112] Yet this requirement is isomorphic with the requirement that the system must start with a highly specified complexity. Thus, to explain the origin of specified biological complexity at the systems level, Kauffman must presuppose the existence of highly specific and complex (i.e., information-rich) molecules as well as a highly specific arrangement of those molecules at the molecular level. Therefore, his work—if it has any relevance to the actual behavior of molecules—presupposes or transfers, rather than explains, the ultimate origin of specified complexity or information.

Others have claimed that the RNA-world scenario offers a promising approach to the origin-of-life problem and with it, presumably, the problem of the origin of the first genetic information. The RNA world was proposed as an explanation for the origin of the interdependence of nucleic acids and proteins in the cell's information-processing system. In extant cells, building proteins requires genetic information from DNA, but information on DNA cannot be processed without many specific proteins and protein complexes. This poses a chicken-or-egg problem. The discovery that RNA (a nucleic acid) possesses some limited catalytic properties similar to those of proteins suggested a way to solve that problem. "RNA-first" advocates proposed an early state in which RNA performed both the enzymatic functions of modern proteins and the information-storage function of modern DNA, thus allegedly making the interdependence of DNA and proteins unnecessary in the earliest living system.

Nevertheless, many fundamental difficulties with the RNA-world scenario have emerged. First, synthesizing (and/or maintaining) many essential building blocks of RNA molecules under realistic conditions has proven either difficult or impossible.[113] Further, the chemical conditions required for the synthesis of ribose sugars are decidedly incompatible with the conditions required for synthesizing nucleotide bases.[114] Yet both are necessary constituents of RNA. Second, naturally occurring RNA possesses very few of the specific enzymatic properties of proteins necessary to extant cells. Third, RNA-world advocates offer no plausible explanation for how primitive RNA replicators might have evolved into modern cells that do rely almost exclusively on proteins to process genetic information and regulate metabolism.[115] Fourth, attempts to enhance the limited catalytic properties of RNA molecules in so-called ribozyme engineering experiments have inevitably required extensive investigator manipulation, thus simulating, if anything, the need for intelligent design, not the efficacy of an undirected chemical evolutionary process.[116]

Most important for our present considerations, the RNA-world hypothesis presupposes, but does not explain, the origin of sequence specificity or information in the original functional RNA molecules. Indeed, the RNA-world scenario was proposed as an explanation for the functional interdependence problem, not the information problem. Even so, some RNA-world advocates seem to envision leapfrogging the sequence-specificity problem. They imagine oligomers of RNA arising by chance on the prebiotic earth and then later acquiring an ability to polymerize copies of themselves—that is, to self-replicate. In such a scenario, the capacity to

self-replicate would favor the survival of those RNA molecules that could do so and would thus favor the specific sequences that the first self-replicating molecules happened to have. Thus, sequences that originally arose by chance would subsequently acquire a functional significance as "an accidental choice remembered."

Like Quastler's DNA-first model, however, this suggestion merely shifts the specificity problem out of view. First, for strands of RNA to perform enzymatic functions (including enzymatically mediated self-replication), they must, like proteins, have very specific arrangements of constituent building blocks (nucleotides in the RNA case). Further, the strands must be long enough to fold into complex three-dimensional shapes (to form so-called tertiary structures). Thus, any RNA molecule capable of enzymatic function must have the properties of complexity and specificity exhibited by DNA and proteins. Hence, such molecules must possess considerable (specified) information content. And yet explaining how the building blocks of RNA might have arranged themselves into functionally specified sequences has proven no easier than explaining how the constituent parts of DNA might have done so, especially given the high probability of destructive cross-reactions between desirable and undesirable molecules in any realistic prebiotic soup. As de Duve has noted in a critique of the RNA-world hypothesis, "hitching the components together in the right manner raises additional problems of such magnitude that no one has yet attempted to do so in a prebiotic context."[117]

Second, for a single-stranded RNA catalyst to self-replicate (the only function that could be selected in a prebiotic environment), it must find another catalytic RNA molecule in close vicinity to function as a template, since a single-stranded RNA cannot function as both enzyme and template. Thus, even if an originally unspecified RNA sequence might later acquire functional significance by chance, it could perform a function only if another RNA molecule—that is, one with a highly specific sequence relative to the original—arose in close vicinity to it. Thus, the attempt to bypass the need for specific sequencing in an original catalytic RNA only shifts the specificity problem elsewhere, namely, to a second and necessarily highly specific RNA sequence. Put differently, in addition to the specificity required to give the first RNA molecule self-replicating capability, a second RNA molecule with an extremely specific sequence—one with essentially the same sequence as the original—would also have to arise. Yet RNA-world theorists do not explain the origin of the requisite specificity in either the original molecule or its twin. Joyce and Orgel have calculated that

to have a reasonable chance of finding two identical RNA molecules of a length sufficient to perform enzymatic functions would require an RNA library of some 10^{54} RNA molecules.[118] The mass of such a library vastly exceeds the mass of the earth, suggesting the extreme implausibility of the chance origin of a primitive replicator system. Yet one cannot invoke natural selection to explain the origin of such primitive replicators, since natural selection only ensues once self-replication has arisen. Further, RNA bases, like DNA bases, do not manifest self-organizational bonding affinities that could explain their specific sequencing. In short, the same kind of evidentiary and theoretical problems emerge whether one proposes that genetic information arose first in RNA or DNA molecules. The attempt to leapfrog the sequencing problem by starting with RNA replicators only shifts the problem to the specific sequences that would make such replication possible.

III.

A. The Return of the Design Hypothesis

If attempts to solve the information problem only relocate it, and if neither chance nor physical-chemical necessity, nor the two acting in combination, explains the ultimate origin of specified biological information, what does? Do we know of any entity that has the causal powers to create large amounts of specified information? We do. As Henry Quastler recognized, the "creation of new information is habitually associated with conscious activity."[119]

Experience affirms that specified complexity or information (defined hereafter as *specified* complexity) routinely arises from the activity of intelligent agents. A computer user who traces the information on a screen back to its source invariably comes to a *mind,* that of a software engineer or programmer. Similarly, the information in a book or newspaper column ultimately derives from a writer—from a mental, rather than a strictly material, cause.

Further, our experience-based knowledge of information-flow confirms that systems with large amounts of specified complexity or information (especially codes and languages) *invariably* originate from an intelligent source—that is, from a mind or a personal agent.[120] Moreover, this generalization holds not only for the semantically specified information present in natural languages but also for other forms of information or specified complexity whether present in machine codes, machines, or works of art.

Like the letters in a section of meaningful text, the parts in a working engine represent a highly improbable yet functionally specified configuration. Similarly, the highly improbable shapes in the rock on Mount Rushmore conform to an independently given pattern: the faces of American presidents known from books and paintings. Thus, both systems have a large amount of specified complexity or information so defined. Not coincidentally, they also originated by intelligent design, not by chance and/or physical-chemical necessity.

This generalization—that intelligence is the only known cause of specified complexity or information (at least, starting from a nonbiological source)—has received support from origin-of-life research itself. During the last forty years, every naturalistic model proposed has failed to explain the origin of the specified genetic information required to build a living cell.[121] Thus, mind or intelligence, or what philosophers call "agent causation," now stands as the only cause known to be capable of generating large amounts of information starting from a nonliving state.[122] As a result, the presence of specified information-rich sequences in even the simplest living systems would seem to imply intelligent design.[123]

Recently, a formal theoretical account of design reasoning has been developed that supports this conclusion. In *The Design Inference*, mathematician and probability theorist William Dembski notes that rational agents often infer or detect the prior activity of other minds by the character of the effects they leave behind. Archaeologists assume, for example, that rational agents produced the inscriptions on the Rosetta stone; insurance-fraud investigators detect certain "cheating patterns" that suggest intentional manipulation of circumstances rather than "natural" disasters; cryptographers distinguish between random signals and those that carry encoded messages. Dembski's work shows that recognizing the activity of intelligent agents constitutes a common and fully rational mode of inference.[124]

More important, Dembski identifies two criteria that typically enable human observers to recognize intelligent activity and to distinguish the effects of such activity from the effects of strictly material causes. He notes that we invariably attribute systems, sequences, or events that have the joint properties of "high complexity" (or low probability) and "specification" (see part I, section E) to intelligent causes—to design—not to chance or physical-chemical laws.[125] By contrast, he notes that we typically attribute to chance those low or intermediate probability events that do not conform to discernable patterns. We attribute to necessity highly probable events that repeatedly recur in a regular or lawlike way.

These inference patterns reflect our knowledge of the way the world works. Since experience teaches, for example, that complex and specified events or systems invariably arise from intelligent causes, we can infer intelligent design from events that exhibit the joint properties of complexity and specificity. Dembski's work suggests a comparative evaluation process for deciding between natural and intelligent causes based on the probabilistic features or "signatures" they leave behind.[126] This evaluation process constitutes, in effect, a scientific method for detecting the activity of intelligence in the echo of its effects.

A homespun example illustrates Dembski's method and criteria of design detection. When visitors first enter Victoria Harbor in Canada from the sea, they notice a hillside awash in red and yellow flowers. As they get closer, they reflexively, and correctly, infer design. Why? Observers quickly recognize a complex and specified pattern, an arrangement of flowers spelling "Welcome to Victoria." They infer the past activity of an intelligent cause—in this case, the careful planning of gardeners. Had the flowers been more haphazardly scattered so as to defy pattern-recognition, observers might have justifiably attributed the arrangement to chance—for example, to random gusts of wind scattering the seed. Had the colors been segregated by elevation, the pattern might have been explained by some natural necessity, such as certain types of plants requiring particular environments or soil types. But since the arrangement exhibits both complexity (the specific arrangement of flowers is highly improbable given the space of possible arrangements) and specificity (the pattern of flowers conforms to the independent requirements of English grammar and vocabulary), observers naturally infer intelligent design. As it turns out, these twin criteria are equivalent (or isomorphic, see part I, section E) with the notion of information as used in molecular biology. Thus, Dembski's theory, when applied to molecular biology, implies that intelligent design played a role in the origin of (specified) biological information.

The logical calculus underlying this inference follows a valid and well-established method used in all historical and forensic sciences. In historical sciences, knowledge of the present causal powers of various entities and processes enables scientists to make inferences about possible causes in the past. When a thorough study of various possible causes turns up only a single adequate cause for a given effect, historical or forensic scientists can make definitive inferences about the past.[127]

The Martian landscape, for example, displays erosional features—trenches and rills—that resemble those produced on Earth by moving

water. Though Mars at present has no significant liquid water on its surface, some planetary scientists have nevertheless inferred that Mars did have a significant amount of water on its surface in the past. Why? Geologists and planetologists have not observed any cause other than moving water that can produce the kind of erosional features that we observe on Mars today. Since in our experience water alone produces erosional trenches and rills, the presence of these features on Mars allows planetologists to infer the past action of water on the surface of the red planet.

Or consider another example. Several years ago one of the forensic pathologists from the original Warren Commission that investigated the assassination of President Kennedy spoke out to quash rumors about a second gunman firing from in front of the motorcade. The bullet hole in the back of President Kennedy's skull apparently evidenced a distinctive beveling pattern that clearly indicated that it had entered his skull from the rear. The pathologist called the beveling pattern a "distinctive diagnostic" because the pattern indicated a single possible direction of entry. Since a rear entry was necessary to cause the beveling pattern in the back of the president's skull, the pattern allowed the forensic pathologists to diagnose the trajectory of the bullet.[128]

Logically, one can infer a cause from its effect (or an antecedent from a consequent) when the cause (or antecedent) is known to be necessary to produce the effect in question. If it's true that "where there's smoke there's fire," then the presence of smoke billowing over a hillside allows us to infer a fire beyond our view. Inferences based on knowledge of empirically necessary conditions or causes ("distinctive diagnostics") are common in historical and forensic sciences and often lead to the detection of intelligent as well as natural causes and events. Since criminal X's fingers are the only known cause of criminal X's fingerprints, X's prints on the murder weapon incriminate him with a high degree of certainty. Similarly, since intelligent design is the only known cause of large amounts of specified complexity or information, the presence of such information implies an intelligent source.

Indeed, since experience affirms mind or intelligent design as a necessary condition (and necessary cause) of information, one can detect (or retrodict) the past action of an intelligence from an information-rich effect—even if the cause itself cannot be directly observed.[129] Thus, the pattern of flowers spelling "Welcome to Victoria" allows visitors to infer the activity of intelligent agents even if they did not see the flowers planted or arranged. Similarly, the specified and complex arrangement of nucleotide

sequences—the information—in DNA implies the past action of an intelligence, even if such mental activity cannot be directly observed.

Scientists in many fields recognize the connection between intelligence and information and make inferences accordingly. Archaeologists assume that a scribe produced the inscriptions on the Rosetta stone; evolutionary anthropologists establish the intelligence of early hominids from chipped flints that are too improbably specified in form (and function) to have been produced by natural causes; NASA's search for extraterrestrial intelligence (SETI) presupposes that any information embedded in electromagnetic signals coming from space would indicate an intelligent source.[130] As yet, however, radio-astronomers have not found any such information-bearing signals. But closer to home, molecular biologists have identified information-rich sequences and systems in the cell, suggesting, by the same logic, an intelligent cause for those effects.

B. Argument from Ignorance? Or Inference to the Best Explanation?

Some would object that any such argument to design constitutes an argument from ignorance. Objectors charge that design advocates use our present ignorance of any sufficient natural cause of information as the sole basis for inferring an intelligent cause of the information present in the cell. Since we don't yet know how biological information could have arisen, we invoke the mysterious notion of intelligent design. On this view, intelligent design functions not as an explanation but as a placeholder for ignorance.

Although the inference to design from the presence of information in DNA does not qualify as a deductively certain proof of intelligent design (empirically based arguments in science rarely do), it does not constitute a fallacious argument from ignorance. Arguments from ignorance occur when evidence against a proposition X is offered as the sole (and conclusive) grounds for accepting some alternative proposition Y.

The inference to design as sketched above (see part III, section A) does not commit this fallacy. True, the previous section of this essay (see part II, sections A-F) argued that at present all types of natural causes and mechanisms fail to account for the origin of biological information from a prebiotic state. And clearly, this lack of knowledge of any adequate natural cause does provide part of the grounds for inferring design from information in the cell. (Though one could just as easily argue that even this "absence of knowledge" actually constitutes a knowledge of absence.) In any

case, our "ignorance" of any sufficient natural cause is only part of the basis inferring design. We also *know* that intelligent agents can and do produce information-rich systems: we have positive experience-based knowledge of an alternative cause that is sufficient, namely, intelligence.

For this reason, the design inference defended here does not constitute an argument from ignorance but an inference to the best explanation.[131] Inferences to the best explanation do not assert the adequacy of one causal explanation merely on the basis of the inadequacy of some other causal explanation. Instead, they compare the explanatory power of many competing hypotheses to determine which hypothesis would, if true, provide the best explanation for some set of relevant data. Recent work on the method of "inference to the best explanation" suggests that determining which among a set of competing possible explanations constitutes the best depends on knowledge of the causal powers of competing explanatory entities.[132]

For example, both an earthquake and a bomb could explain the destruction of the building, but only a bomb could explain the presence of charring and shrapnel at the scene of the rubble. Earthquakes do not produce shrapnel, nor do they cause charring, at least not on their own. Thus, the bomb best explains the pattern of destruction at the building site. Entities, conditions, or processes that have the capability (or causal powers) to produce the evidence in question constitute better explanations of that evidence than those that do not.

It follows that the process of determining the best explanation often involves generating a list of possible hypotheses, comparing their known (or theoretically plausible) causal powers with respect to the relevant data, and then progressively eliminating potential but inadequate explanations, and finally, in the best case, electing the one remaining causally adequate explanation.

This essay has followed precisely this method to make a case for intelligent design as the best explanation for the origin of biological information. It has evaluated and compared the causal efficacy of four broad categories of explanation—chance, necessity, the combination of those two, and intelligent design—with respect to their ability to produce large amounts of specified complexity or information. As we have seen, neither scenarios based on chance nor those based on necessity (nor those that combine the two) can explain the origin of specified biological information in a prebiotic context. That result comports with our uniform human experience. Natural processes do not produce information-rich structures starting from

purely physical or chemical antecedents. Nor does matter, whether acting at random or under the force of physical-chemical necessity, arrange itself into complex, information-rich sequences.

Nevertheless, it is not correct to say that we do not know how information arises. We know from experience that conscious intelligent agents can create informational sequences and systems. To quote Quastler again, the "creation of new information is habitually associated with conscious activity."[133] Further, experience teaches that whenever large amounts of specified complexity or information are present in an artifact or entity whose causal story is known, invariably creative intelligence—intelligent design—played a causal role in the origin of that entity. Thus, when we encounter such information in the biomacromolecules necessary to life, we may infer—based on our *knowledge* of established cause-effect relationships—that an intelligent cause operated in the past to produce the specified complexity or information necessary to the origin of life.

As formulated, this inference to design employs the same method of argumentation and reasoning that historical scientists use generally. Indeed, in the *Origin of Species,* Darwin himself developed his argument for universal common ancestry as an inference to the best explanation. As he explained in a letter to Asa Gray:

> I . . . test this hypothesis [common descent] by comparison with as many general and pretty well-established propositions as I can find—in geographical distribution, geological history, affinities &c., &c. And it seems to me that, *supposing* that such a hypothesis were *to explain* such general propositions, we ought, in accordance with the common way of following all sciences, to admit it till some *better* hypothesis be found out [emphasis added].[134]

Moreover, as formulated, the argument to design from the information in DNA also adheres to the standard uniformitarian canons of method employed within the historical sciences. The principle of uniformitarianism states that "the present is the key to the past." In particular, the principle specifies that our knowledge of present cause-effect relationships should govern our assessment of the plausibility of the inferences that we make about the remote causal past. Yet it is precisely such knowledge of cause-effect relationships that informs the inference to intelligent design. Since we know that intelligent agents do produce large amounts of information, and since all known natural processes do not (or cannot), we can infer design as the best explanation of the origin of information in the cell. Recent

developments in the information sciences (such as Dembski's work in *The Design Inference*) help to define and formalize knowledge of such cause-effect relationships, allowing us to make inferences about the causal histories of various artifacts, entities, or events based on the complexity and information-theoretic signatures they exhibit.[135]

In any case, the inference to design depends on present *knowledge* of the demonstrated causal powers of natural entities and intelligent agency, respectively. It no more constitutes an argument from ignorance than any other well-grounded inference in geology, archaeology, or paleontology—where present knowledge of cause-effect relationships guides the inferences that scientists make about the causal past.

Objectors might still deny the legitimacy of inferring intelligent design (even as a best explanation) because we are ignorant of what future inquiry may uncover about the causal powers of other materialistic entities or processes. Some would characterize the design inference presented here as invalid or unscientific because it depends on a negative generalization—that is, "purely physical and chemical causes do not generate large amounts of specified information"—which future discoveries may later falsify. We should "never say never," they say.

Yet science often says "never," even if it can't say so for sure. Negative or proscriptive generalizations often play an important role in science. As many scientists and philosophers of science have pointed out, scientific laws often tell us not only what does happen but also what does not happen.[136] The conservation laws in thermodynamics, for example, proscribe certain outcomes. The first law tells us that energy is never created or destroyed. The second tells us that the entropy of a closed system will never decrease over time. Those who claim that such "proscriptive laws" do not constitute *knowledge,* because they are based on past but not future experience, will not get very far if they try to use their skepticism to justify funding for research on, say, perpetual motion machines.

Further, without proscriptive generalizations, without knowledge about what various possible causes cannot or do not produce, historical scientists could not make determinations about the past. Reconstructing the past requires making abductive inferences from present effects back to past causal events.[137] Making such inferences requires a progressive elimination of competing causal hypotheses. Deciding which causes can be eliminated from consideration requires knowing what effects a given cause can—and cannot—produce. If historical scientists could never say that particular entities lack particular causal powers, they could never eliminate them, even

provisionally, from consideration. Thus, they could never infer that a specific cause had acted in the past. Yet historical and forensic scientists make such inferences all the time.

Moreover, Dembski's examples of design inferences—from fields such as archaeology, cryptography, fraud-detection, and criminal forensics—show that we often infer the past activity of an intelligent cause and do so, evidently, without worrying about committing fallacious arguments from ignorance. And we do so for good reason. A vast amount of human experience shows that intelligent agents have unique causal powers that matter (especially nonliving matter) does not. When we observe features or effects that we know from experience only agents produce, we rightly infer the prior activity of intelligence.

To determine the best explanation, scientists do not need to say "never" with absolute certainty. They need only say that a postulated cause is best, given what we know at present about the demonstrated causal powers of competing entities or agencies. That cause C can produce effect E makes it a better explanation of E than some cause D that has never produced E (especially if D seems incapable of doing so on theoretical grounds), even if D might later demonstrate causal powers of which we are presently ignorant.[138]

Thus, the objection that the design inference constitutes an argument from ignorance reduces in essence to a restatement of the problem of induction. Yet one could make the same objection against any scientific law or explanation or against any historical inference that takes present, but not future, knowledge of natural laws and causal powers into account. As Barrow and Tipler have noted, to criticize design arguments, as Hume did, simply because they assume the uniformity and (normative character) of natural law cuts just as deeply against "the rational basis of any form of scientific inquiry."[139] Our knowledge of what can and cannot produce large amounts of specified information may later have to be revised, but so might the laws of thermodynamics. Inferences to design may later prove incorrect, as may other inferences implicating various natural causes. Such possibilities do not stop scientists from making generalizations about the causal powers of various entities or from using those generalizations to identify probable or most plausible causes in particular cases.

Inferences based on past and present experience constitute knowledge (albeit provisional), not ignorance. Those who object to such inferences object to *science* as much as they object to a particular science-based hypothesis of design.

C. But Is It Science?

Of course, many simply refuse to consider the design hypothesis on grounds that it does not qualify as "scientific." Such critics affirm an extra-evidential principle known as methodological naturalism.[140] Methodological naturalism asserts that, as a matter of definition, for a hypothesis, theory, or explanation to qualify as "scientific," it must invoke only naturalistic or materialistic entities. On that definition, critics say, the intelligent design hypothesis does not qualify. Yet, even if one grants this definition, it does not follow that some nonscientific (as defined by methodological naturalism) or metaphysical hypothesis may not constitute a better, more causally adequate, explanation. This essay has argued that, whatever its classification, the design hypothesis does constitute a better explanation than its materialistic or naturalistic rivals for the origin of specified biological information. Surely, simply classifying an argument as metaphysical does not refute it.

In any case, methodological naturalism now lacks justification as a normative definition of science. First, attempts to justify methodological naturalism by reference to metaphysically neutral (that is, non-question-begging) demarcation criteria have failed.[141] Second, to assert methodological naturalism as a normative principle for all of science has a negative effect on the practice of certain scientific disciplines, especially the historical sciences. In origin-of-life research, for example, methodological naturalism artificially restricts inquiry and prevents scientists from seeking some hypotheses that might provide the best, most causally adequate explanations. To be a truth-seeking endeavor, the question that origin-of-life research must address is not "Which materialistic scenario seems most adequate?" but rather "What actually caused life to arise on Earth?" Clearly, one possible answer to that latter question is this one: "Life was designed by an intelligent agent that existed before the advent of humans." If one accepts methodological naturalism as normative, however, scientists may never consider the design hypothesis as possibly true. Such an exclusionary logic diminishes the significance of any claim of theoretical superiority for any remaining hypothesis and raises the possibility that the best "scientific" explanation (as defined by methodological naturalism) may not be the best in fact.

As many historians and philosophers of science now recognize, scientific theory-evaluation is an inherently comparative enterprise. Theories that gain acceptance in artificially constrained competitions can claim to be neither "most probably true" nor "most empirically adequate." At best, such

theories can be considered the "most probably true or adequate among an artificially limited set of options." Openness to the design hypothesis would seem necessary, therefore, to any fully rational historical biology—that is, to one that seeks the truth, "no holds barred."[142] A historical biology committed to following the evidence wherever it leads will not exclude hypotheses a priori on metaphysical grounds. Instead, it will employ only metaphysically neutral criteria—such as explanatory power and causal adequacy—to evaluate competing hypotheses. Yet this more open (and seemingly rational) approach to scientific theory evaluation would now suggest the theory of intelligent design as the best, most causally adequate, explanation for the origin of the information necessary to build the first living organism.

Notes

1. Harmke Kamminga, "Protoplasm and the Gene," in *Clay Minerals and the Origin of Life,* ed. A. G. Cairns-Smith and H. Hartman (Cambridge: Cambridge University Press, 1986), 1.
2. Alexander Oparin, *Genesis and Evolutionary Development of Life* (New York: Academic Press, 1968), 7.
3. F. Crick and J. Watson, "A Structure for Deoxyribose Nucleic Acid," *Nature* 171 (1953): 737–38; F. Crick and J. Watson, "Genetical Implications of the Structure of Deoxyribose Nucleic Acid," *Nature* 171 (1953): 964–67, esp. 964; T. D. Schneider, "Information Content of Individual Genetic Sequences," *Journal of Theoretical Biology* 189 (1997): 427–41; W. R. Loewenstein, *The Touchstone of Life: Molecular Information, Cell Communication, and the Foundations of Life* (New York: Oxford University Press, 1999).
4. Bernd-Olaf Kuppers, *Information and the Origin of Life* (Cambridge: MIT Press, 1990), 170–72.
5. L. E. Kay, "Who Wrote the Book of Life? Information and the Transformation of Molecular Biology," *Science in Context* 8 (1994): 601–34; L. E. Kay, "Cybernetics, Information, Life: The Emergence of Scriptural Representations of Heredity," *Configurations* 5 (1999): 23–91; L. E. Kay, *Who Wrote the Book of Life?* (Stanford, Calif.: Stanford University Press, 2000), xv–xix.
6. Darwin's only speculation on the origin of life is found in an unpublished 1871 letter to Joseph Hooker. In it, he sketched the outlines of the chemical evolutionary idea, namely, that life could have first

evolved from a series of chemical reactions. As he envisioned it, "if (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc., that a proteine compound was chemically formed ready to undergo still more complex changes." Cambridge University Library, Manuscripts Room, Darwin Archives, courtesy Peter Gautrey.

7. E. Haeckel, *The Wonders of Life,* trans. J. McCabe (London: Watts, 1905), 111; T. H. Huxley, "On the Physical Basis of Life," *Fortnightly Review* 5 (1869): 129–45.
8. A. I. Oparin, *The Origin of Life,* trans. S. Morgulis (New York: Macmillan, 1938); S. C. Meyer, "Of Clues and Causes: A Methodological Interpretation of Origin of Life Studies" (Ph.D. diss., Cambridge University, 1991).
9. W. T. Astbury and A. Street, "X-Ray Studies of the Structure of Hair, Wool and Related Fibers," *Philosophical Transactions of the Royal Society of London* A 230 (1932): 75–101; H. Judson, *Eighth Day of Creation* (New York: Simon and Schuster, 1979), 80; R. Olby, *The Path to the Double Helix* (London: Macmillan, 1974), 63.
10. Olby, *Path to the Double Helix,* 7, 265.
11. Judson, *Eighth Day,* 213, 229–35, 255–61, 304, 334–35, 562–63; F. Sanger and E. O. P. Thompson, "The Amino Acid Sequence in the Glycyl Chain of Insulin," parts 1 and 2, *Biochemical Journal* 53 (1953): 353–66, 366–74.
12. Judson, *Eighth Day,* 562–63; J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, and H. Wyckoff, "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis," *Nature* 181 (1958): 662–66, esp. 664.
13. B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell* (New York: Garland, 1983), 111–12, 127–31.
14. Judson, *Eighth Day,* 30.
15. Ibid., 30–31, 33–41, 609–10; Oswald T. Avery, C. M. MacCleod, and M. McCarty, "Induction of Transformation by a Deoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III," *Journal of Experimental Medicine* 79 (1944): 137–58.
16. Judson, *Eighth Day,* 95–96; E. Chargaff, *Essays on Nucleic Acids* (Amsterdam: Elsevier, 1963), 21.
17. Chargaff, *Essays,* 21.
18. Crick and Watson, "Structure."

19. Judson, *Eighth Day,* 96.
20. Crick and Watson, "Genetical Implications," 964–67.
21. Judson, *Eighth Day,* 611.
22. Crick and Watson, "Structure"; Crick and Watson, "Genetical Implications."
23. Judson, *Eighth Day,* 245–46, 335–36.
24. Ibid., 470–89; J. H. Matthei and M. W. Nirenberg, "Characteristics and Stabilization of DNAase-Sensitive Protein Synthesis in *E. coli* Extracts," *Proceedings of the National Academy of Sciences, USA* 47 (1961): 1580–88; J. H. Matthei and M. W. Nirenberg, "The Dependence of Cell-Free Protein Synthesis in *E. coli* upon Naturally Occurring or Synthetic Polyribonucleotides," *Proceedings of the National Academy of Sciences, USA* 47 (1961): 1588–1602.
25. Alberts et al., *Molecular Biology,* 106–8; S. L. Wolfe, *Molecular and Cellular Biology* (Belmont, Calif.: Wadsworth, 1993), 639–48.
26. We now know, of course, that in addition to the process of gene expression, specific enzymes must often modify amino acid chains after translation in order to achieve the precise sequencing necessary to allow correct folding into a functional protein. The amino acid chains produced by gene expression may also undergo further modification in sequence at the endoplasmic reticulum. Finally, even well-modified amino acid chains may require preexisting protein "chaperons" to help them fold into a functional three-dimensional configuration. All these factors make it impossible to predict a protein's final sequence from its corresponding gene sequence alone. See S. Sarkar, "Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology," in *The Philosophy and History of Molecular Biology: New Perspectives,* ed. S. Sarkar (Dordrecht, Netherlands: Boston Studies in Philosophy of Science, 1996), 196, 199–202. Nevertheless, this unpredictability in no way undermines the claim that DNA exhibits the property of "sequence specificity," or the isomorphic claim that it contains "specified information" as argued in part I, section E. Sarkar argues, for example, that the absence of such predictability renders the concept of information theoretically superfluous for molecular biology. Instead, this unpredictability shows that the sequence specificity of DNA base sequences constitutes a necessary, though not sufficient, condition of attaining proper protein folding—that is, DNA does contain specified information (part I, section E), but not enough to determine protein folding by itself. Instead, the presence of both

post-translation processes of modification and pretranscriptional genomic editing (through exonucleases, endonucleases, spliceosomes, and other editing enzymes) only underscores the need for other preexisting, information-rich biomolecules in order to process genomic information in the cell. The presence of a complex and functionally integrated information-processing system *does* suggest that the information on the DNA molecule is insufficient to produce proteins. It does not show that such information is *unnecessary* to produce proteins, nor does it invalidate the claim that DNA, therefore, stores and transmits specified genetic information.

27. C. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* 27 (1948): 379–423, 623–56.
28. F. Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981), 6–10.
29. Ibid.; Shannon, "A Mathematical Theory."
30. B. Kuppers, "On the Prior Probability of the Existence of Life," in *The Probabilistic Revolution,* ed. Lorenz Kruger et al. (Cambridge: MIT Press, 1987), 355–69.
31. Schneider, "Information Content"; see also H. P. Yockey, *Information Theory and Molecular Biology* (Cambridge: Cambridge University Press, 1992), 246–58, for important refinements in the method of calculating the information-carrying capacity of proteins and DNA.
32. C. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Urbana: University of Illinois Press, 1949), 8.
33. Schneider, "Information Content," 58–177; Yockey, *Information Theory,* 58–177.
34. See note 26. Sarkar, "Biological Information," 199–202, esp. 196; F. Crick, "On Protein Synthesis," *Symposium for the Society of Experimental Biology* 12 (1958): 138–63, esp. 144, 153.
35. Crick, "On Protein Synthesis," 144, 153.
36. Recall that the determination of the genetic code depended, for example, on observed correlations between changes in nucleotide base sequences and amino acid production in "cell-free systems." See Judson, *Eighth Day,* 470–87.
37. W. A. Dembski, *The Design Inference: Eliminating Chance Through Small Probabilities* (Cambridge: Cambridge University Press, 1998), 1–35, 136–74.
38. Ibid., 136–74.

39. Of the two sequences, only the second meets an independent set of functional requirements. To convey meaning in English one must employ preexisting (or independent) conventions of vocabulary (associations of symbol sequences with particular objects, concepts, or ideas) and existing conventions of syntax and grammar (such as "every sentence requires a subject and a verb"). When arrangements of symbols "match" or utilize these vocabulary and grammatical conventions (that is, functional requirements), meaningful communication can occur in English. The second sequence ("Time and tide wait for no man") clearly exhibits such a match between itself and preexisting requirements of vocabulary and grammar. The second sequence has employed these conventions to express a meaningful idea. It also, therefore, falls within the smaller (and conditionally independent) pattern delimiting the domain of all meaningful sentences in English and thus, again, exhibits a "specification."
40. J. Bowie and R. Sauer, "Identifying Determinants of Folding and Activity for a Protein of Unknown Sequences: Tolerance to Amino Acid Substitution," *Proceedings of the National Academy of Sciences, USA* 86 (1989): 2152–56; J. Reidhaar-Olson and R. Sauer, "Functionally Acceptable Solutions in Two Alpha-Helical Regions of Lambda Repressor," *Proteins, Structure, Function, and Genetics* 7 (1990): 306–10.
41. R. Dawkins, *River out of Eden* (New York: Basic Books, 1995), 11.
42. B. Gates, *The Road Ahead* (Boulder, Colo.: Blue Penguin, 1996), 228.
43. L. E. Orgel, *The Origins of Life on Earth* (New York: John Wiley, 1973), 189.
44. See note 5. Kay, "Who Wrote," 611–12, 629; Kay, "Cybernetics"; Kay, *Who Wrote.*
45. Sarkar, "Biological Information," 199–202.
46. E. Schrödinger, *What Is Life? And Mind and Matter* (Cambridge: Cambridge University Press, 1967), 82; Alberts et al., *Molecular Biology,* 21; Crick and Watson, "A Structure"; Crick and Watson, "Genetical Implications"; Crick, "On Protein"; Judson, *Eighth Day,* 611; Orgel, *Origins of Life,* 189.
47. P. Davies, *The Fifth Miracle* (New York: Simon and Schuster, 1998), 120.
48. Orgel, *Origins of Life,* 189.
49. Loewenstein, *Touchstone;* Davies, *Fifth Miracle;* Schneider, "Information Content"; C. Thaxton and W. Bradley, "Information and the Origin of Life," in *The Creation Hypothesis: Scientific Evidence for an Intelligent*

Designer, ed. J. P. Moreland (Downers Grove, Ill.: InterVarsity Press, 1994), 173–210, esp. 190; S. Kauffman, *The Origins of Order* (Oxford: Oxford University Press, 1993), 287–340; Yockey, *Information Theory,* 178–293; Kuppers, *Information and Origin,* 170–72; F. Crick, *Life Itself* (New York: Simon and Schuster, 1981), 59–60, 88; J. Monod, *Chance and Necessity* (New York: Vintage Books, 1971), 97–98, 143; Orgel, *Origins,* 189; D. Kenyon and G. Steinman, *Biochemical Predestination* (New York: McGraw-Hill, 1969), 199–211, 263–66; Oparin, *Genesis,* 146–47; H. Quastler, *The Emergence of Biological Organization* (New Haven, Conn.: Yale University Press, 1964).

50. G. Wald, "The Origin of Life," *Scientific American* 191 (August 1954): 44–53; R. Shapiro, *Origins: A Skeptic's Guide to the Creation of Life on Earth* (New York: Summit Books, 1986), 121.
51. F. Crick, "The Origin of the Genetic Code," *Journal of Molecular Biology* 38 (1968): 367–79; H. Kamminga, "Studies in the History of Ideas on the Origin of Life" (Ph.D. diss., University of London 1980), 303–4.
52. C. de Duve, "The Constraints of Chance," *Scientific American* (Jan. 1996): 112; Crick, *Life Itself,* 89–93; Quastler, *Emergence,* 7.
53. H. J. Morowitz, *Energy Flow in Biology* (New York: Academic Press, 1968), 5–12; F. Hoyle and C. Wickramasinghe, *Evolution from Space* (London: J. M. Dent, 1981), 24–27; A. G. Cairns-Smith, *The Life Puzzle* (Edinburgh: Oliver and Boyd, 1971), 91–96; I. Prigogine, G. Nicolis, and A. Babloyantz, "Thermodynamics of Evolution," *Physics Today* (23 Nov. 1972); Yockey, *Information Theory,* 246–58; H. P. Yockey, "Self-Organization, Origin of Life Scenarios and Information Theory," *Journal of Theoretical Biology* 91 (1981): 13–31; Bowie and Sauer, "Identifying Determinants"; Reidhaar-Olson et al., *Proteins;* Shapiro, *Origins,* 117–31.
54. Prigogine, "Thermodynamics."
55. Cairns-Smith, *Life Puzzle,* 95.
56. Reidhaar-Olson and Sauer, "Functionally Acceptable"; D. D. Axe, "Biological Function Places Unexpectedly Tight Constraints on Protein Sequences," *Journal of Molecular Biology* 301, no. 3: 585–96; M. Behe, "Experimental Support for Regarding Functional Classes of Proteins to Be Highly Isolated from Each Other," in *Darwinism: Science or Philosophy?* ed. J. Buell and V. Hearn (Richardson, Tex.: Foundation for Thought and Ethics, 1994), 60–71; Yockey, *Information Theory,* 246–58. Actually, Sauer counted sequences that folded into stable three-dimensional configurations as functional, though many sequences

that fold are not functional. Thus, his results actually underestimate the probabilistic difficulty.

57. Behe, "Experimental Support."
58. Axe, "Biological Function."
59. Dawkins, *Blind Watchmaker,* 54, 139.
60. Dembski, *Design Inference,* 175–223; E. Borel, *Probabilities and Life,* trans. M. Baudin (New York: Dover, 1962), 28. Dembski's universal probability bound actually reflects the "specificational" resources, not the probabilistic resources in the universe. Dembski's calculation determines the number of specifications possible in finite time. It nevertheless has the effect of limiting the "probabilistic resources" available to explain the origin of any *specified* event of small probability. Since living systems are precisely specified systems of small probability, the universal probability bound effectively limits the probabilistic resources available to explain the origin of specified biological information.
61. Dembski, *Design Inference,* 175–223. Cassette mutagenesis experiments have usually been performed on proteins of about 100 amino acids in length. Yet extrapolations from these results can generate reasonable estimates for the improbability of longer protein molecules. For example, Sauer's results on the proteins lambda repressor and arc repressor suggest that, on average, the probability at each site of finding an amino acid that will maintain functional sequencing (or, more accurately, that will produce folding) is less than 1 in 4 (1 in 4.4). Multiplying 1/4 by itself 150 times (for a protein 150 amino acids in length) yields a probability of roughly 1 chance in 10^{91}. For a protein of that length, the probability of attaining both exclusive peptide bonding and homochirality is also about 1 chance in 10^{91}. Thus, the probability of achieving all the necessary conditions of function for a protein 150 amino acids in length exceeds 1 chance in 10^{180}.
62. Dembski, *Design Inference,* 67–91, 175–214; Borel, *Probabilities,* 28.
63. E. Pennisi, "Seeking Life's Bare Genetic Necessities," *Science* 272 (1996): 1098–99; A. Mushegian and E. Koonin, "A Minimal Gene Set for Cellular Life Derived by Comparison of Complete Bacterial Genomes," *Proceedings of the National Academy of Sciences, USA* 93 (1996): 10268–73; C. Bult et al., "Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschi,*" *Science* 273 (1996): 1058–72.
64. Dembski, *Design Inference,* 67–91, 175–223, 209–10.

65. P. T. Mora, "Urge and Molecular Biology," *Nature* 199 (1963): 212–19.
66. I. Hacking, *The Logic of Statistical Inference* (Cambridge: Cambridge University Press, 1965), 74–75.
67. Dembski, *Design Inference,* 47–55.
68. C. de Duve, "The Beginnings of Life on Earth," *American Scientist* 83 (1995): 437.
69. Quastler, *Emergence,* 7.
70. Oparin, *Origin of Life,* 64–103; Meyer, *Of Clues,* 174–79, 194–98, 211–12.
71. Oparin, *Origin of Life,* 107–8, 133–35, 148–59, 195–96.
72. Oparin, *Genesis,* 146–47.
73. C. de Duve, *Blueprint for a Cell: The Nature and Origin of Life* (Burlington, N.C.: Neil Patterson, 1991), 187.
74. G. Joyce and L. Orgel, "Prospects for Understanding the Origin of the RNA World," in *RNA World,* ed. R. F. Gesteland and J. J. Atkins (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press, 1993), 1–25, esp. 8–13.
75. T. Dobzhansky, "Discussion of G. Schramm's Paper," in *The Origins of Prebiological Systems and of Their Molecular Matrices,* ed. S. W. Fox (New York: Academic Press, 1965), 310; H. H. Pattee, "The Problem of Biological Hierarchy," in *Toward a Theoretical Biology,* ed. C. H. Waddington, vol. 3 (Edinburgh: Edinburgh University Press, 1970), 123.
76. P. T. Mora, "The Folly of Probability," in Fox, *Origins,* 311–12; L. V. Bertalanffy, *Robots, Men and Minds* (New York: George Braziller, 1967), 82.
77. J. Von Neumann, *Theory of Self-reproducing Automata,* completed and edited by A. Berks (Urbana: University of Illinois Press, 1966).
78. Pennisi, "Seeking"; Mushegian and Koonin, "Minimal Gene Set"; Bult et al., "Complete Genome Sequence."
79. E. Wigner, "The Probability of the Existence of a Self-reproducing Unit," in *The Logic of Personal Knowledge,* ed. E. Shils (London: Kegan and Paul, 1961), 231–35; P. T. Landsberg, "Does Quantum Mechanics Exclude Life?" *Nature* 203 (1964): 928–30; H. J. Morowitz, "The Minimum Size of the Cell," in *Principles of Biomolecular Organization,* ed. M. O'Connor and G. E. W. Wolstenholme (London: J. A. Churchill, 1966), 446–59; Morowitz, *Energy Flow,* 10–11.
80. Dawkins, *Blind Watchmaker,* 47–49; Kuppers, "On the Prior Probability."
81. Kuppers, "On the Prior Probability," 366.

82. Dawkins, *Blind Watchmaker,* 47–49; P. Nelson, "Anatomy of a Still-Born Analogy," *Origins and Design* 17 (3) (1996): 12.
83. de Duve, "Beginnings of Life," 437.
84. Morowitz, *Energy Flow,* 5–12.
85. G. Steinman and M. N. Cole, "Synthesis of Biologically Pertinent Peptides Under Possible Primordial Conditions," *Proceedings of the National Academy of Sciences, USA* 58 (1967): 735–41; G. Steinman, "Sequence Generation in Prebiological Peptide Synthesis," *Archives of Biochemistry and Biophysics* 121 (1967): 533–39; R. A. Kok, J. A. Taylor, and W. L. Bradley, "A Statistical Examination of Self-Ordering of Amino Acids in Proteins," *Origins of Life and Evolution of the Biosphere* 18 (1988): 135–42.
86. Kenyon and Steinman, *Biochemical Predestination,* 199–211, 263–66.
87. I. Prigogine and G. Nicolis, *Self-Organization in NonEquilibrium Systems* (New York: John Wiley, 1977), 339–53, 429–47.
88. Kauffman, *Origins of Order,* 285–341; de Duve, "Beginnings of Life"; C. de Duve, *Vital Dust: Life as a Cosmic Imperative* (New York: Basic Books, 1995).
89. C. Thaxton, W. Bradley, and R. Olsen, *The Mystery of Life's Origin: Reassessing Current Theories* (Dallas: Lewis and Stanley, 1992), v–viii; D. Kenyon and G. Mills, "The RNA World: A Critique," *Origins and Design* 17, no. 1 (1996): 9–16; D. Kenyon and P. W. Davis, *Of Pandas and People: The Central Question of Biological Origins* (Dallas: Haughton, 1993); S. C. Meyer, "A Scopes Trial for the '90's," *Wall Street Journal,* 6 Dec. 1993; Kok et al., "Statistical Examination."
90. Steinman and Cole, "Synthesis"; Steinman, "Sequence Generation."
91. Kok et al., "Statistical Examination"; B.J. Strait and G. T. Dewey, "The Shannon Information Entropy of Biologically Pertinent Peptides," *Biophysical Journal* 71: 148–155.
92. Kuppers, "On the Prior Probability," 64.
93. de Duve, "Beginnings of Life," 437.
94. R. Stalnaker, *Inquiry* (Cambridge: MIT Press, 1984), 85.
95. This, in fact, happens where adenine and thymine do interact chemically in the complementary base-pairing *across* the information-bearing axis of the DNA molecule. *Along* the message bearing axis, however, there are no chemical bonds or differential bonding affinities that determine sequencing.
96. M. Polanyi, "Life's Irreducible Structure," *Science* 160 (1968): 1308–12, esp. 1309.

97. As noted in part I, section D, the information-carrying capacity of any symbol in a sequence is inversely related to the probability of its occurrence. The informational capacity of a sequence as a whole is inversely proportional to the product of the individual probabilities of each member in the sequence. Since chemical affinities between constituents ("symbols") increase the probability of the occurrence of one, given another (i.e., necessity increases probability), such affinities decrease the information-carrying capacity of a system in proportion to the strength and relative frequency of such affinities within the system.
98. Dretske, *Knowledge and the Flow,* 12.
99. Yockey, "Self-Organization," 18.
100. H. P. Yockey, "A Calculation of the Probability of Spontaneous Biogenesis by Information Theory," *Journal of Theoretical Biology* 67 (1977): 377–98, esp. 380.
101. M. Eigen, *Steps Toward Life* (Oxford: Oxford University Press, 1992), 12.
102. Quastler, *Emergence,* ix.
103. Ibid., 1, 47.
104. Yockey, *Information Theory,* 247.
105. J. C. Walton, "Organization and the Origin of Life," *Origins* 4 (1977): 16–35.
106. J. M. Smith, "Hypercycles and the Origin of Life," *Nature* 280 (1979): 445–46; F. Dyson, *Origins of Life* (Cambridge: Cambridge University Press, 1985), 9–11, 35–39, 65–66, 78; Shapiro, *Origins,* 161.
107. Iris Fry, *The Emergence of Life on Earth* (New Brunswick, N.J.: Rutgers University Press, 2000), 158.
108. Kauffman, *Origins of Order,* 285–341.
109. Ibid., 299.
110. See Protein Databank at http://www.rcsb.org/pdb.
111. Kauffman, *Origins of Order,* 298.
112. Thaxton, et al., *Mystery of Life's Origin,* 127–43.
113. R. Shapiro, "Prebiotic Cytosine Synthesis: A Critical Analysis and Implications for the Origin of Life," *Proceedings of the National Academy of Sciences, USA* 96 (1999): 4396–4401; M. M. Waldrop, "Did Life Really Start Out in an RNA World?" *Science* 246 (1989): 1248–49.
114. R. Shapiro, "Prebiotic Ribose Synthesis: A Critical Analysis," *Origins of Life and Evolution of the Biosphere* 18 (1988): 71–85; Kenyon and Mills, "RNA World."

115. G. F. Joyce, "RNA Evolution and the Origins of Life," *Nature* 338 (1989): 217–24.
116. A. J. Hager, J. D. Polland Jr., and J. W. Szostak, "Ribozymes: Aiming at RNA Replication and Protein Synthesis," *Chemistry and Biology* 3 (1996): 717–25.
117. de Duve, *Vital Dust,* 23.
118. Joyce and Orgel, "Prospects for Understanding," 1–25, esp. 11.
119. Quastler, *Emergence,* 16.
120. A possible exception to this generalization might occur in biological evolution. If the Darwinian mechanism of natural selection acting on random variation can account for the emergence of all complex life, then a mechanism does exist that can produce large amounts of information—assuming, of course, a large amount of *preexisting* biological information in a self-replicating living system. Thus, even if one assumes that the selection/variation mechanism can produce all the information required for the macroevolution of complex life from simpler life, that mechanism will not suffice to account for the origin of the information necessary to produce life from nonliving chemicals. As we have seen, appeals to *prebiotic* natural selection only beg the question of the origin of specified information. Thus, based on experience, we can affirm the following generalization: "for all nonbiological systems, large amounts [note 118 below] of specified complexity or information originate only from mental agency, conscious activity, or intelligent design." Strictly speaking, *experience* may even affirm a less qualified generalization (such as "large amounts of specified invariably originate from an intelligent source"), since the claim that natural selection acting on random mutations can produce large amounts of novel genetic information depends on debatable theoretical arguments and extrapolation from observations of small-scale microevolutionary changes that do not themselves manifest large gains in biological information. Later in this volume (in "The Cambrian Explosion: Biology's Big Bang"), Meyer, Ross, Nelson, and Chien argue that neither the neo-Darwinian mechanism nor any other current naturalistic mechanism adequately accounts for the origin of the information required to build the novel proteins and body plans that arise in the Cambrian explosion. In any case, the more qualified empirical generalization (stated above in this endnote) is sufficient to support the argument presented here, since this essay seeks only to establish intelligent design as the best explanation for

origin of the specified information necessary to the origin of the *first* life.

121. K. Dose, "The Origin of Life: More Questions Than Answers," *Interdisciplinary Science Reviews* 13 (1988): 348–56; Yockey, *Information Theory,* 259–93; Thaxton et al., *Mystery,* 42–172; Thaxton and Bradley, "Information and the Origin," 193–97; Shapiro, *Origins.*
122. Of course, the phrase "large amounts of specified information" again begs a quantitative question, namely, "How much specified information or complexity would the minimally complex cell have to have before it implied design?" Recall that Dembski has calculated a universal probability bound of $1/10^{150}$ corresponding to the probabilistic/specificational resources of the known universe. Recall further that probability is inversely related to information by a logarithmic function. Thus, the universal small probability bound of $1/10^{150}$ translates into roughly 500 bits of information. Chance alone, therefore, does not constitute a sufficient explanation for the de novo origin of any specified sequence or system containing more than 500 bits of (specified) information. Further, since systems characterized by complexity (a lack of redundant order) defy explanation by self-organizational laws and since appeals to prebiotic natural selection presuppose but do not explain the origin of the specified information necessary to a minimally complex self-replicating system, intelligent design best explains the origin of the more than 500 bits of specified information required to produce the first minimally complex living system. Thus, assuming a nonbiological starting point (note 116 above), the de novo emergence of 500 or more bits of specified information will reliably indicate design.
123. Again, this claim applies at least in cases where the competing causal entities or conditions are nonbiological—or where the mechanism of natural selection can be safely eliminated as an inadequate means of producing requisite specified information.
124. Dembski, *Design Inference,* 1–35.
125. Ibid., 1–35, 136–223.
126. Ibid., 36–66.
127. Ibid.; E. Sober, *Reconstructing the Past* (Cambridge, Mass.: MIT Press, 1988), 4–5; M. Scriven, "Causes, Connections, and Conditions in History," in *Philosophical Analysis and History,* ed. W. Dray (New York: Harper and Row, 1966), 238–64, esp. 249–50.
128. *McNeil-Lehrer News Hour,* Transcript 19 (May 1992).

129. Meyer, *Of Clues,* 77–140.
130. Less exotic (and more successful) design-detection occurs routinely in both science and industry. Fraud-detection, forensic science, and cryptography all depend on the application of probabilistic or information theoretic criteria of intelligent design. Dembski, *Design Inference,* 1–35. Many would admit that we *may* justifiably infer a past human intelligence operating (within the scope of human history) from an information-rich artifact or event, but only because we already know that human minds exist. But, they argue, since we do not know whether an intelligent agent(s) existed prior to humans, inferring the action of a designing agent that antedates humans cannot be justified, even if we observe an information-rich effect. Note, however, that SETI scientists do not already know whether an extraterrestrial intelligence exists. Yet they assume that the presence of a large amount of specified information (such as the first 100 prime numbers in sequence) would definitively establish the existence of one. Indeed, SETI seeks precisely to establish the existence of other intelligences in an unknown domain. Similarly, anthropologists have often revised their estimates for the beginning of human history or civilization because they discovered information-rich artifacts dating from times that antedate their previous estimates. Most inferences to design establish the existence or activity of a mental agent operating in a time or place where the presence of such agency was previously unknown. Thus, to infer the activity of a designing intelligence from a time prior to the advent of humans on Earth does not have a qualitatively different epistemological status than other design inferences that critics already accept as legitimate. T. R. McDonough, *The Search for Extraterrestrial Intelligence: Listening for Life in the Cosmos* (New York: Wiley, 1987).
131. P. Lipton, *Inference to the Best Explanation* (New York: Routledge, 1991), 32–88.
132. Ibid.; S. C. Meyer, "The Scientific Status of Intelligent Design: The Methodological Equivalence of Naturalistic and Non-Naturalistic Origins Theories," in *Science and Evidence for Design in the Universe, The Proceedings of the Wethersfield Institute,* vol. 9 (San Francisco: Ignatius Press, 2000), 151–212; Meyer, "The Demarcation of Science and Religion," in *The History of Science and Religion in the Western Tradition: An Encyclopedia,* ed. G. B. Ferngren (New York: Garland, 2000), 17–23; E. Sober,

The Philosophy of Biology (San Francisco: Westview Press, 1993); Meyer, *Of Clues,* 77–140.

133. Quastler, *Emergence,* 16.
134. Francis Darwin, ed., *Life and Letters of Charles Darwin,* 2 vols. (London: D. Appleton, 1896), 1:437.
135. Dembski, *Design Inference,* 36–37, esp. 37.
136. Oparin, *Origin of Life,* 28; M. Rothman, *The Science Gap* (Buffalo, N.Y.: Prometheus, 1992), 65–92; K. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (London: Routledge and Kegan Paul, 1962), 35–37.
137. Meyer, *Of Clues,* 77–140; Sober, *Reconstructing the Past,* 4–5; de Duve, "Beginnings of Life," 249–50.
138. R. Harre and E. H. Madden, *Causal Powers* (London: Basil Blackwell, 1975).
139. J. Barrow and F. Tipler, *The Anthropic Cosmological Principle* (Oxford: Oxford University Press, 1986), 69.
140. M. Ruse, "*McLean v. Arkansas:* Witness Testimony Sheet," in *But Is It Science?* ed. M. Ruse (Amherst, N.Y.: Prometheus Books, 1988), 103; Meyer, "Scientific Status"; Meyer, "Demarcation."
141. Meyer, "Scientific Status"; Meyer, "Demarcation"; L. Laudan, "The Demise of the Demarcation Problem," in Ruse, *But Is It Science?* 337–50; L. Laudan, "Science at the Bar—Causes for Concern," in Ruse, *But Is It Science?* 351–55; A. Plantinga, "Methodological Naturalism?" *Origins and Design* 18, no. 1 (1986): 18–26; A. Plantinga, "Methodological Naturalism?" *Origins and Design* 18, no. 2 (1986): 22–34.
142. Bridgman, *Reflections of a Physicist,* 2d ed. (New York: Philosophical Library, 1955), 535.