God and the World of Signs: Semiotics and the Emergence of Life

with Andrew Robinson and Christopher Southgate, "Introduction: Toward a Metaphysic of Meaning"; Christopher Southgate and Andrew Robinson, "Interpretation and the Origin of Life"; Bruce H. Weber, "Selection, Interpretation, and the Emergence of Living Systems"; Jesper Hoffmeyer, "A Biosemiotic Approach to the Question of Meaning"; Robert E. Ulanowicz, "Process Ecology: Stepping Stones to Biosemiosis"; Andrew Robinson and Christopher Southgate with Terrence Deacon, "Discussion of the Conceptual Basis of Biosemiotics"

INTERPRETATION AND THE ORIGIN OF LIFE

by Christopher Southgate and Andrew Robinson

Abstract. We offer a general definition of interpretation based on a naturalized teleology. The definition tests and extends the biosemiotic paradigm by seeking to provide a philosophically robust resource for investigating the possible role of semiosis (processes of representation and interpretation) in biological systems. We show that our definition provides a way of understanding various possible kinds of misinterpretation, illustrate the definition using examples at the cellular and subcellular level, and test the definition by applying it to a potential counterexample. We explain how we propose to use the definition as a way of asking new questions about what distinguishes life from non-life and of formulating testable hypotheses within the field of origin-of-life research. If the definition leads to fruitful new empirical approaches to the scientific problem of the origin of life, it will help to establish biosemiotics as a legitimate philosophical approach in theoretical biology and will thereby support a theological appropriation of the biosemiotic perspective as the basis of a new theology of nature.

Keywords: autocell; biosemiotics; function; information; interpretation; life; natural selection; origin of life; purpose; ribozyme; semiotics; teleology

---

Christopher Southgate and Andrew Robinson are Hon. University Fellows in the Department of Theology, University of Exeter. Address correspondence to Dr. C.C.B. Southgate, Amory Building, Rennes Drive, Exeter, EX4 4RJ, U.K.; e-mail c.c.b.southgate@ex.ac.uk.

[*Zygon*, vol. 45, no. 2 (June 2010)]

www.zygonjournal.org

We propose that semiosis, and more specifically the capacity of an entity to interpret signs within its environment, is a fundamental property of life. Further, we propose that interpretation may be one helpful diagnostic property (among others) of protobiotic entities and suggest experimental work by which this proposal may be tested. These proposals form part of a wider project involving explorations in the philosophy of biology (Robinson and Southgate 2010b) and in trinitarian theology. The formulation of this new program in the theology of nature and novel scientific predictions arising from such theological reflection are in Robinson and Southgate forthcoming a, b, c; Robinson in press.

In our general introduction to this series of articles on the previous pages (Robinson and Southgate 2010a, 339–44) we spelled out our wider theological project of using semiotics as the basis of a new theology of nature. This project would be significantly weakened if it could be shown that the application of semiotic terminology to fundamental biological processes is a convenient metaphorical usage rather than a reflection of an underlying reality about biology (see Oyama, Griffiths, and Gray 2001, 5). Conversely, if we could show that semiotic concepts are indispensable in theoretical biology, the philosophical and scientific basis of our semiotic theology of nature would be strengthened. In order to investigate whether semiosis is a fundamental property of living things it is necessary to develop a robust general definition of interpretation that shows why interpretative responses cannot be explained merely in terms of mechanistic causation.

INTERPRETATION DEFINED

We begin by setting out a philosophically robust general definition of interpretation that (a) does not presume the concepts it seeks to test and (b) is not simply reducible to mechanistic causal relations. Our definition is a modification of the one offered by T. L. Short in his important work on the semiotics of C. S. Peirce (Short 2004; 2007). Short argues that Charles Darwin's concept of natural selection points to a naturalized and yet irreducible account of purpose (Short 2002). He further shows how Peirce appropriated this naturalized teleology as the basis of his teleological understanding of interpretation (Short 2004). A central element of Short's argument is that purposes are always possible general types of outcome, an insight that can be traced to Aristotle but has become obscured in most contemporary discussions of teleology. The explanatory role of the general type in question is what makes teleological explanations irreducible to mechanistic explanations, because the latter always involve the explanation of particulars by reference to particulars.[1] Thus, a feature of an organism that has been selected by natural selection may be said to "exist for a purpose" because selection, including natural selection, is always selection for a possible general type of outcome. The particular details of the heart have been selected by natural selection because they have a general type of effect, namely, that they facilitate the pumping of blood. There are numerous ways that an organism might evolve a capacity to circulate body fluid. The explanation of the existence of the particular features of the heart cannot be given only by reference to another particular or set of particulars but must include reference to a general—that is, the general property of pumping blood (Short 2007, 111). Something can "exist for a purpose" or "have a purpose" by virtue of a process of selection that is itself purposeless, such as natural selection.

In general, Short's account of purpose may be expressed as follows:

A purpose *P* of something *Q* is:

> a general type of outcome or effect *P*
>
> which explains the existence of *Q*,
>
> because *Q* is selected for its having effects of general type *P*
>
> (where *P* is not constitutively related to *Q*).[2]

Our definition of interpretation is a modification of Short's proposal for a definition of interpretation, intended to ensure that the definition presupposes none of the concepts that it is intended to explain. Only by doing so will it be suitable for application to biological processes and, specifically, to our proposed approach to the emergence of life. Drawing on Short's (Peircean) account of purposes as general types of outcome we therefore have proposed the following definition (Robinson and Southgate 2010b):

A response, *R*, of an entity is an interpretant of some *X* as a sign of some object *O* if and only if:

1. The entity has a property, *Q*, of undergoing change of state *S* in response to some *X*, where *R* is any actual instance of such a response;
2. (a) *R* tends to increase the probability of an effect of a certain general type, *P*;

   (b) This tendency of *R* depends on a relation between *X* and *O*, where the occurrence of *X* does not necessarily imply the occurrence of *O*;
3. The property *Q* has been selected for the tendency of instances of *R* to actualize effects of general type *P*.

See Figure 1 for diagrammatic representation of the definition.

The definition may be illustrated by the following hypothetical scenario of a "hungry" amoeba. Suppose that *X* is a gradient of some chemical attractant; *Q* is the property of the amoeba of responding to the attractant gradient by changing state (*S*) to the state of crawling up the gradient; *R* is a particular instance of the amoeba's responding to an attractant gradient by changing state (*S*) to swim up the gradient; *O* is the presence of a bacterium in the direction of the attractant gradient; and *P* is the outcome of the amoeba's "having a meal" (that is, the general possible type of effect of obtaining nutrition by ingesting a bacterium).

According to our definition, *R* interprets *X* as a sign of *O* because:

1. The amoeba has a property, *Q*, of responding to something *X* (attractant gradient) by undergoing change of state *S* (starting to crawl in the direction of the attractant gradient), where *R* is any actual instance of such a response;
2. (a) *R* (responding to attractant gradient by crawling up the gradient) tends to increase the probability of an effect of a certain general type, *P* (the amoeba's having a meal);

   (b) This tendency of *R* depends on a (fallible) relation between *X* (an attractant gradient) and *O* (the presence of a bacterium);
3. The property *Q* (the propensity for the crawling response) has been selected for the tendency of instances of *R* (instances of the crawling response) to actualize effects of general type *P* (the amoeba's having a meal).

Note that we use the terms *S* and *Q* in order to distinguish clearly between the potential to undergo a particular change of state (*S*), the property (*Q*) of undergoing the change of state as a response to *X*, and actual instances of such a response (*R*). One reason that these distinctions are important is that, considered in isolation, individual instances of response *R* may be merely automatic and mechanical. This point is particularly important when considering what might be the simplest sort of entity capable of making an interpretation. What makes interpretation irreducible to mechanistic explanation is that *Q*, the property of making such responses, has been selected for its having (via particular instances of *R*, even if *R* is itself mechanical) effects of a certain general possible type, *P*. Only when the change of state occurs in response to something, *X*, and all the other criteria in the definition relating *Q*, *S*, *R*, *X*, *O*, and *P* are met does a

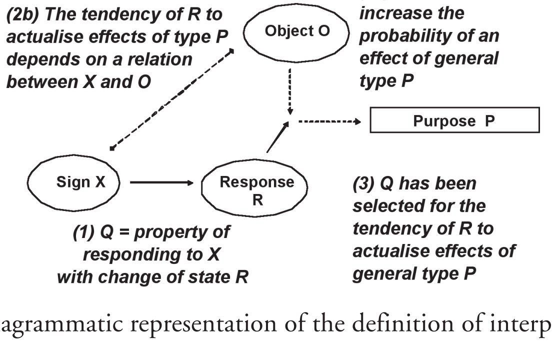

Fig. 1. Diagrammatic representation of the definition of interpretation.

"change of state for a purpose" count as an interpretation. In the simplest scenarios, including our example of the hungry amoeba, the mechanism of selection of the property *Q* will have been Darwinian natural selection. At other "levels" of interpretative response other mechanisms of selection will come into play, such as the deliberate selection of responses by conscious agents.[3]

TYPES OF MISINTERPRETATION

It is intrinsic to such an account of interpretation that the response *R* is a fallible response. Otherwise the definition falls, either because *X* necessarily implies *O* or because *Q* automatically implies *P*. Our definition enables us to clarify different ways in which an entity can make a misinterpretation.

It is helpful to visualize these different ways by considering again the "hungry" amoeba. If the amoeba (otherwise functioning normally) does not respond to the attractant gradient by crawling up that gradient, the property *Q* is "deficient." We therefore term this a "*Q*-level error." A *Q*-level error would occur also if the amoeba, capable of responding correctly to the attractant *X*, responded in the same way to a gradient of the related chemical *X*′, which was not and could not be a sign of the presence of a bacterium. *Q*-level errors, then, occur when the response, although selected for, lacks the sensitivity or specificity to enable the entity to recognize and respond appropriately to the whole range of stimuli (of this particular type) to which it is exposed.[4]

A different case arises when, in a particular instance, there is no bacterium in the direction of the attractant gradient. The general correlation between the sign *X* (the attractant gradient) and the object *O* (presence of a bacterium) may be very strong, but it does not hold in this case. The amoeba responds by crawling up the gradient, but there is no bacterium to crawl after. We term this an "*R*-level error," because it relates to the particular individual instance of response *R*, not to *Q*, the general property of responding. An *R*-level error occurs when the response, although selected for, and indeed constituting a correct recognition of the sign, fails in an individual instance because the sign-object relation does not hold.

In summary, whereas *R*-level misinterpretations occur at the level of individual instances of *R* (clause 2 in our definition), *Q*-level errors occur at the level of selection of the property of responsiveness to *X* (clause 3).

More difficult cases arise when the bacterium is too far away for there to be a net gain in energy for the amoeba from crawling to ingest it, or where the bacterium is indigestible or even toxic. The amoeba may correctly interpret the attractant gradient as a sign of the presence of a bacterium by crawling up the gradient and thereby getting closer to the bacterium. Yet in terms of the ultimate purpose of having a meal this particular response proves to be an error.[5] This raises the complex issue of how exactly the general type of outcome *P* should be described in a case where there is a chain of outcomes from achieving proximity to a bacterium to gaining nutrition from ingesting it. We analyze this issue elsewhere (Robinson and Southgate 2010b).

EXAMPLES OF INTERPRETATION IN A RANGE OF BIOLOGICAL SYSTEMS

We gave the example of the hungry amoeba as an illustration of the application of our definition of interpretation to nonhuman instances of interpretative responsiveness. It may help to clarify the concept we are proposing to give a range of instances in known biological systems of differing complexities. An interesting example in a single-celled organism is the response to photons in certain species of *Euglena*. Such organisms have a "spot" very sensitive to light. A few photons are sufficient to trigger the response of moving toward the light. The tendency of instances of this motile response is to actualize effects of a general type, namely, that the *Euglena* is able to generate considerable amounts of internal metabolic energy (by bathing its chloroplast—a different structure from the light-sensitive spot—with light). Central to this example is the distinction between the response *R* to one or a few photons hitting the "spot," and the prospective benefit—the photosynthetic production of many molecules of ATP (from many photons hitting the chloroplast) that would be required to more than compensate for the energy required to move in the direction of the incoming light.[6] This example also helps us to see that a biological interpretation is like a wager; the property *Q* (of which the response *R* is a specific instance) is selected for because the probability that the response will be, on balance, successful has in the past exceeded the probability that it is not. This generates a probability matrix of evolutionary strategies involving interpretation. This we intend to model in further work.

The examples of the chemotactic amoeba and photosensitive *Euglena* concerned motility in single-celled organisms. The phenomenon of interpretation may also be clearly illustrated in processes within higher organisms. The release of the hormone insulin by the mammalian pancreas has evolved to be exquisitely sensitive to rises in blood glucose. The hormonal system has evolved such that a small rise in glucose, detected by the beta cells of the pancreatic islets of Langerhans, is taken as a sign of a more general condition: that nutrition is being taken in and that it will (probably) be favorable to switch on the anabolic processes by which glucose is converted into metabolic stores (glycogen and fat).

In terms of our definition:

1. The pancreas has a property, *Q*, of responding to something *X* (presence of glucose molecules) by undergoing change of state *S* (secretion of insulin), where *R* is any actual instance of such a response;

2. (a) *R* (responding to the presence of glucose by secreting insulin) tends to increase the probability of an effect of a certain general type, *P* (laying down of absorbed nutrients in metabolic stores);

   (b) This tendency of *R* depends on a (fallible) relation between *X* (the presence of glucose) and *O* (the presence of significant levels of absorbed nutrients);
3. The property *Q* (secretion of insulin in response to the presence of glucose) has been selected for the tendency of instances of *R* (the secretion response) to actualize effects of general type *P* (laying down of absorbed nutrients in metabolic stores).

Another very familiar example from the world of higher organisms is that of neurotransmission. The presence of neurotransmitter molecules in the synaptic cleft is interpreted by the receptors in the "receiving" terminal as a sign that the "donor" axon has fired.[7] Note once again that this interpretation may be incorrect but is sufficiently reliable to make this mechanism of nerve conduction an extremely widespread and evolutionarily effective strategy.[8]

So much is familiar ground and reinforces the point that our definition establishes a continuity between events at the cellular level and interpretations by conscious agents, without in any way minimizing the differences between the processes involved. Consider now a fine example of subcellular biological interpretation in the recently discovered mechanism of the riboswitch. These sophisticated mechanisms for the regulation of the expression of mRNA operate at their most basic in a way that need not be regarded as interpretative. For instance, the so-called glm riboswitch recognizes glucosamine 6-phosphate, and the effect of the recognition is to inhibit the conversion of fructose 6-phosphate to glucosamine 6-phosphate. This is a straightforward metabolic feedback, not interpretative in our terms because there is no sense in which some entity is being taken as a sign of some other entity or condition. However, according to Maumita Mandal and Ronald R. Breaker, the array of S-adenosylmethionine (SAM) riboswitches in an organism such as *B. Anthracis* not only regulates SAM metabolism but allows SAM to act as

> a master molecule, the concentration of which is monitored by numerous riboswitch variants in each cell. Specifically, a drop of SAM concentration might be indicative of a general lack of sulphur-containing compounds. This is independently detected by each of the mRNAs in the regulon, which results in an increase in the expression of genes that are required for sulphur metabolism. (Mandal and Breaker 2004, 457)

Thus, levels of SAM detected at the riboswitches are being taken as a sign of a more general condition. Here we see, in what are widely presumed to be metabolically ancient subcellular systems,[9] that what molecular biology describes meets our definition of interpretation.

COUNTEREXAMPLE: DOES A THERMOSTAT MAKE INTERPRETATIONS?

One way of attempting to refute our definition of interpretation would be to devise thought experiments in which systems to which one would not wish to ascribe interpretative properties may be shown to meet our definition. Consider one such possible counterexample. Suppose that a thermostat, the basis of whose function is a bimetallic strip, switches a room heater off if the room temperature rises above 20°C. First consider the case in which a bimetallic strip, in isolation, had formed by chance by some natural process. Clause 1 of our definition is fulfilled: The strip has a property, *Q*, of undergoing change of state *S* (bending) in response to something, *X* (temperature above 20°C), where *R* is any actual instance of such a response. The temperature responsiveness of this naturally occurring bimetallic strip would not comply with our definition because the responsiveness has not been selected for any tendency of the response, *R*, to produce effects of a general type, *P* (clause 3). Clause 3 would, in contrast, be fulfilled when the strip (whether naturally occurring or designed) is deliberately incorporated as part of a functioning thermostat. In that case the responsiveness of the strip has been selected for the general type of effect, *P*, of keeping the room temperature below 20°C. However, even then the definition of interpretation would not be complied with because the strip responds directly to the room temperature, so there is no disjunction between *X* and *O* that would allow for a fallible relation between the two (clause 2b). Suppose, though, that the thermostat is situated in a draughty corner of the room. The temperature in that corner has a relationship to, but is not always a precise indicator of, the temperature in the rest of the room (so *X* does not necessarily imply *O*, fulfilling clause 2b). Nevertheless, for reasons of convenience or some other practical constraint, the owner of the thermostat still chooses to locate the device in that corner. We might further suppose that the general type of effect (*P*) with which he is concerned is the health of his house plants (so that *O*, the temperature of the room, is not identical to *P*). We concede that with these additional postulates the response of the thermostat meets the definition of interpretation. Importantly, though, it does so only because it functions within a wider context contrived by a conscious purposeful agent. In that context the thermostat may reasonably be held to be performing an interpretative job that the human agent prefers not to have to do himself.[10]

The example of the bimetallic strip thus shows that our definition of interpretation excludes certain instances that would normally be regarded as merely mechanistic effects but allows such effects to count as interpretations if occurring within a wider, tightly defined, context. In the case of the thermostat, that context is the product of a conscious agent.

SIGNIFICANCE FOR ORIGIN-OF-LIFE RESEARCH

We have shown that fallible interpretation (as distinct from straightforward mechanistic feedback mechanisms) is a familiar feature of a whole range of biological systems. We further propose that interpretation may be considered a necessary, though not a sufficient, condition for life. It is hard to imagine a living organism that did not possess properties *Q* of responding to signs *X* of features *O* of their environment by undergoing changes of state, each *Q* having been selected for its tendency to further outcomes of general type *P* (ultimately the furtherance of the reproductive fitness of the organism). Equally, it is hard to imagine an entity that one would not consider to be at least protobiotic that did possess this interpretative capacity. Thomas Sebeok suggested that "a full understanding of the dynamics of semiosis . . . may, in the last analysis, turn out to be no less than the *definition* of life" (1972, 118). We prefer to say that interpretation is a neglected diagnostic feature for assessing the likelihood that a proposed protobiotic system may have led to the emergence of life.

We agree, then, with those who suggest that several minimal criteria for the emergence of life are likely to be important, none constituting in isolation a definition of life. Stuart Kauffman and Philip Clayton list five such criteria: (1) autocatalytic reproduction, (2) the operation of a work cycle, (3) a boundary to individuate the entity, (4) the construction of constraints to the release of work, and (5) "choice and action that have evolved to respond to (e.g.) food or poison" (Kauffman and Clayton 2006, 500). Properties corresponding to each of these criteria must have emerged in some sequence or other in the systems within which life emerged. Model systems can be proposed that emphasize each type of property. One reason why origin-of-life research remains somewhat polarized is that workers in the field tend to espouse the particular virtues of their own model, be it a model based on "replication first" or "metabolism first" or "work-cycle first" or "localization first."[11] Further work on such models would benefit from not only concentrating on developing the integrity and internal coherence of each type of model in itself. Rather, in considering whether a model system is in plausible continuity with both the simpler conditions within which it is supposed to have emerged and more lifelike systems to which it may have given rise, it may prove important to consider the interactions between the different criteria. We hypothesize that the acquisition of a capacity for making interpretations of the state of the environment, internal or external, would have given rise to a selective advantage over similar protobiotic entities lacking such a capacity. Origin-of-life research therefore would benefit from asking for each of the approaches (replication-first, metabolism-first, enclosure-first) what might be the role of selection for an interpretative capacity as a possible next evolutionary step in the emergence of life.

In our thinking, all of the criteria discussed above for promising protolife are taken to precede the existence of "representational information" of the sort that DNA is often considered to constitute in a modern cell. Much confusion has arisen in the literature through assuming that such information or coding must be the earliest, or even an early, step in the emergence of life, or that such representation is the minimum criterion for semiosis. (See Küppers 1990 for a typical study; also Schneider 2000). Insisting on the importance of coding tends to mean invoking a complex nucleic acid–protein interaction at an early stage; rejection of coding often has meant the rejection of all early roles for nucleic acid. Our definition of interpretation allows for the possibility of an early role for nucleic acid sequences without imputing to these any property of coding.

We propose two specific empirical tests of the application of our definition of interpretation to the problem of the origin of life. The first involves computer modeling of Terrence Deacon's "autocell," a hypothetical protobiotic system consisting of a simple autocatalytic set of molecules (Deacon 2006). We have developed a computer model of the synthesis of non-interpretative and interpretative autocells under kinetic control in systems in which substrates arrive simultaneously in pulses, followed by intervals of low or zero substrate concentration. The modeling has demonstrated that autocells whose dissociation is catalyzed by a molecule *X*, the presence of which is positively correlated with pulses of substrate, will, under appropriate conditions, be synthesized at a higher rate than those whose dissociation occurs merely randomly. These *X*-sensitive autocells are, within the terms of our definition, interpreting a molecule of *X* as a sign of the presence of favorable conditions in which to dissociate. We further predict that if *X* is sometimes present in the absence of substrates, the interpretative autocells will misinterpret the presence of *X* as a sign of favorable conditions (an *R*-level error), resulting in a reduced rate of synthesis. Preliminary results have illustrated both the advantage of interpretation and the disadvantage of misinterpretation in terms of the net production of autocells in this model system (Lui et al. 2010).

The autocell is a hypothetical protobiotic system involving an autocatalytic set with a minimum of four substrates and two catalytic products. We suggest that even simpler molecular systems may be capable of interpreting their environments. Specifically, we have proposed that a single molecule that can exist in two conformations, and that has a propensity to change state in response to a particular environmental circumstance, may be regarded as the simplest interpretative entity (Robinson and Southgate 2010b). We refer to this hypothetical molecule as a "minimal molecular semiotic entity."

Our second empirical test of the application of the definition of interpretation involves the possibility of synthesizing an RNA enzyme (ribozyme) with such a property. This could be achieved by starting with a ribozyme that catalyzes the ligation of two half-molecules to make more of itself. This catalysis is always metal ion-dependent, and the most common metal ion to be supplied is Mg^{2+}. Using an "artificial evolution" process (see Joyce 2004), it would then be possible to select positively for variants that bind an unusual metal ion such as Mn^{2+} and are effective catalysts of the ligation of their subcomponents at (say) pH6. This would be followed by negative selection for variants of the Mn^{2+} dependent enzyme that are inactivated at pH8.5 in the presence of the metal ion. In the resulting ribozyme, the presence of Mn^{2+} would result in a conformational change to bind the ion. This new conformation—open to the binding of substrates in a configuration appropriate to their ligation—would, if the two half-molecules were present and the pH in the optimal range for catalysis, result in enhanced production of more copies of the ribozyme. However, if the binding of Mn^{2+} took place at high pH, the ribozyme, being in the "open" conformation, would have an increased tendency to be degraded. If the environment fluctuates between favorable conditions (low pH, presence of substrate) and unfavorable conditions (high pH, absence of substrate), and if there is a correlation between, but not an invariable association of, the presence of Mn^{2+} and favorable conditions, this artificially evolved ribozyme will be capable of making an interpretation as follows:

1. The ribozyme has a property, *Q*, of responding to something *X* (presence of Mn^{2+}) by undergoing change of state *S* (conformational change), where *R* is any actual instance of such a response;
2. (a) *R* (responding to Mn^{2+} by undergoing a conformational change) tends to increase the probability of an effect of a certain general type, *P* (production of more ribozyme);

   (b) This tendency of *R* depends on a (fallible) relation between *X* (presence of Mn^{2+}) and *O* (the presence of substrate in favorable conditions for replication);
3. The property *Q* (the propensity for this conformational change) has been selected for the tendency of instances of *R* (instances of the conformational change in response to Mn^{2+}) to actualize effects of general type *P* (production of more ribozyme).

Once such a molecule had been "engineered" it would be possible to allow the nucleotide sequence of the ribozyme to evolve during further iterations of the artificial-evolution process using different variations of "environmental" conditions and variable associations of the concentration of the Mn^{2+} cofactor (the sign, *X*) with favorable conditions of low pH and high substrate concentration (object *O*). We predict that for any set of evolutionary conditions with a given probabilistic relation between *X* and *O* the interpretative ribozyme would evolve toward optimization of the sensitivity to *X* so as to balance the benefit of responding to *X* in favorable circumstances against the disadvantage of responding to *X* in unfavorable circumstances. Furthermore, once such a molecule had been "engineered" and tested, the properties it manifested would provide criteria for investigating naturally occurring ribozymes for similar properties. This would thus potentially provide a way of linking currently occurring RNA molecules with the hypothetical ancient "RNA world" (Gesteland, Cech, and Atkins 2006).

We recognize, of course, that any RNA catalyst found in contemporary organisms has undergone billions of years of biochemical evolution, yet the very persistence of such catalysts (most strikingly at the locus of peptide formation in the ribosome) testifies to their metabolic importance and to the extent of the selection pressure to retain them relatively unaltered.[12] We also recognize the difficulties associated with emphasizing the centrality of RNA in prebiotic scenarios and stress that what we are attempting is not yet another effort to short-circuit "the molecular biologist's dream" (Orgel 2002, 142).[13] Nevertheless, if RNA sequences were important in the emergence of life, not necessarily as replicons but in the first instance as a heterogeneous population of oligomers constituting an autocatalytic set, the characterization of interpretation in naturally occurring RNA catalysts, in continuity with the properties of engineered ribozymes and subcellular arrays of riboswitches, would be an important indicator that this property may have contributed significantly to the evolutionary survival and indeed success of such molecules.

In addition to our investigation of ribozymes as possible minimal molecular semiotic entities we consider it important to apply the search for interpretation across a wide range of systems in which protobiotic properties are being postulated. For this reason we propose to investigate whether the permeability of spontaneously forming vesicles could also have been a locus of prebiotic interpretation. Here the testable scenarios are less clear at present. One interesting line of work is the investigation of the association of small ionophoric peptides into complexes that create pores in biological membranes (Pohorille et al. 2008). It may be possible to demonstrate that "deengineered" versions of such peptides, in which a key residue was missing, could bind the missing amino acid and associate into pore-forming complexes. On that basis it may be possible to characterize the properties of a mechanism in which recognition of that amino acid led to a vesicle's becoming much more permeable to the substrates of a putative autocatalytic set (perhaps including an energy-transducing system) contained within its interior.

INTERPRETATION AS A PROBABLE NEXT EVOLUTIONARY STEP IN ORIGIN-OF-LIFE SCENARIOS

We are now in a position to summarize the relevance of our thinking to the field of origin-of-life research in general. As noted, origin-of-life work often focuses on one of several possible necessary criteria for the emergence of life. We suggest that for each of these criteria—such as localization, autocatalysis, replication, or the performance of a work cycle—it is useful to ask what would be the "probable next evolutionary step" (PNES) for the system in question. Thinking about the problem of the origin of life in this way helps us to focus on the interactions between the key elements of the different, currently competing, proposals. This strategy is illustrated in Table 1. Origin-of-life proposals are listed in the left column. Subsequent columns indicate (with a) whether one or more of various functional attributes

TABLE 1
Comparison of Origin-of-Life Proposals

Key Properties	Interpretation	Localization	Autocatalysis	Thermo-dynamic work-cycle
Model System				
Robinson-Southgate minimal molecular semiotic entity	√		PNES	PNES[a]
Kauffman hexamer-trimer system[b]	√[c]	PNES[d]	√	√
Deacon autocell[e]	PNES[f]	√	√	
Vesicle system[g]	PNES[h]	√	√	√[i]
Kauffman-Clayton minimal autonomous agent[j]	√	√	√	√

Notes to Table 1. (PNES = probable next evolutionary step)

a. Acquisition of the capacity to catalyze the formation of its own activated monomers using an external free energy source and (for example) pyrophosphate as high-energy intermediate (see Kauffman 2000). The first steps of such evolution would count first as autocatalysis, then leading to the formation of a work cycle.

b. Kauffman 2000.

c. We contend that the allosteric feed-forward and feedback mechanisms in this model do in fact provide a good example of interpretative faculties hidden within Kauffman's model.

d. As suggested in Kauffman and Clayton 2006.

e. Deacon 2006.

f. As suggested in Robinson and Southgate 2010b.

g. In for example the work of Deamer 1997.

h. Perhaps through pore formation in the vesicle in response to first contact with external supplies of potential substrates for the internal autocatalytic set.

i. D. W. Deamer discovered that certain types of container could develop a proton gradient in response to light (Deamer 1997). Such a gradient could be the ancient precursor of biochemical chemiosmosis and could form the basis of a work-cycle.

j. Kauffman and Clayton 2006. That the Kauffman-Clayton minimal autonomous agent possesses all these properties is not surprising given the sophistication they regard the system as having to possess in order to manifest a "choosing" capability.

(top row) are central to the proposal in question. For each of the proposals we also indicate what we consider to be the PNES.

The Table reflects our conviction, for example, that the Deacon autocell would be likely to acquire interpretation as a PNES (Robinson and Southgate 2010b). In contrast, our minimal molecular semiotic entity, already possessing the property of interpretation, would be likely to acquire a property such as a work cycle, providing the energy for its replication, or an autocatalytic cycle.

Our analysis thus provides an additional diagnostic property of protobiotic entities, that of capacity for making interpretative responses, which we commend to researchers working on the emergence of life. We acknowledge that attributing a property of interpretation (though not agency or choice; Robinson and Southgate 2010b) to relatively simple molecules may seem counterintuitive. For this reason some may prefer to refer to the most minimal examples of changes of state that meet our definition as protointerpretations rather than interpretations. Whichever term one uses, the important point is that our definition picks out a phenomenon that has a formal continuity with undisputed instances of interpretation and that is irreducible to purely mechanistic explanation.

CONCLUSION

C. S. Peirce's account of the triadic, teleological, and fallible nature of semiotic processes can be developed into a tightly formulated general definition of interpretation capable of covering all instances and levels of interpretation. Such a definition of interpretation accounts for various possible types of misinterpretation and can be shown to fit a variety of processes occurring at cellular and subcellular levels of biological organization. Our definition suggests new empirical approaches to the scientific problem of the origin of life. If the definition proves to be fruitful in this regard, as preliminary results suggest, it would support the philosophical program of biosemiotics. If biosemiotic thinking can be shown to generate testable and fruitful scientific hypotheses, it will be justifiable to appropriate the biosemiotic perspective as the basis of a theology of nature.[14]

NOTES

We are grateful to Dr. T. L. Short for a generous exchange of views and various helpful personal communications. We thank Drs. Philip Clayton, Terrence Deacon, Niels Gregersen, Jesper Hoffmeyer, Stuart Kauffman, Niles Lehman, Robert Ulanowicz, Bruce Weber, and Mark Wynn for very helpful consultations on this material. We are very glad to acknowledge the support of the Science and Transcendence Advanced Research Series 2007, 2008, and 2009, an initiative of the Center for Theology and the Natural Sciences, Berkeley, California, with funding from the John Templeton Foundation.

1. This is not to deny that a conception of generals is required as a prerequisite for mechanistic explanations, only to assert that the explanans of mechanistic explanations is not a general type of possible outcome.

2. The element of the definition in parentheses is intended to exclude "self-organizing" systems from the definition of purpose. For example, Bénard cells (*Q*) exist because in certain circumstances that organization of molecular motions has the general type of effect (*P*) of dissipating energy. This general type of effect, *P*, explains the selection of *Q* from among the other possible patterns of molecular motion that might arise randomly. But we would not normally say that dissipation of energy is the "purpose" of a Bénard cell because *P* is constitutively related to *Q*. In other words, in systems that are only self-organizing, there is no distinction between means and ends.

3. Because our definition relies on the property *Q* having been selected for, it has the following somewhat paradoxical implication in cases of natural selection (rather than selection by a conscious agent): The first time the response *R* takes place, it cannot constitute an interpretation, even if it is an advantageous response to *X* being a sign of *O*. On that first occasion *R* does not occur for a purpose and cannot be "correct" or "mistaken."

4. Strictly speaking, a (mis)interpretative response based on a *Q*-level error does not in itself meet our definition of interpretation because the property *Q* in such instances cannot comply with clause 3. We have shown elsewhere (by making a distinction, not discussed here, between "proper" and "non-proper" interpretation) that *Q*-level errors can nevertheless be understood in the light of our definition (Robinson and Southgate 2010b).

5. It would be an *R*-level error if the attractant gradient were exactly the same as in the case where the bacterium was accessible and digestible. If the inaccessible, indigestible, or toxic bacterium gave rise to a different attractant gradient, which the amoeba was unable to distinguish from that of the accessible and digestible bacterium, the response would be a *Q*-level error.

6. We thank Dr. David Deamer for this example.

7. An even simpler mechanism of cell-to-cell communication is found in the cellular slime mould, where low nutrient levels lead to the secretion of cyclic AMP (Goodwin 1995, 44–49).

8. Neural interpretation is of course the prime mechanism making possible conscious interpretation in higher organisms. But in introducing it as an example we are not thinking particularly of such complex networks; extremely simple neural systems exhibit the same phenomenon. An important form of "misinterpretation" now exploited clinically is the case where reuptake of the transmitter has been inhibited. The interpretation that presence of the transmitter implies the recent firing of the donor axon is then incorrect. The "tricking" of the serotoninergic systems in the human brain in this way is the basis of the mechanism of a whole class of antidepressants.

9. As Wade C. Winkler and Ronald R. Breaker note,

> protein is most likely the superior medium for forming such structures. In other words, a protein will most likely be tapped by a cell if a new metabolite-sensing factor is needed for gene control purposes. If this is true, then perhaps many of the riboswitches we see in modern cells are the distant relatives of metabolite-binding RNAs that might have been present before the evolutionary emergence of proteins. Given that many riboswitch ligands appear to be ancient in origin, this possibility cannot be ruled out. Also, these ligands have not changed over the last several billion years of evolution, and so the RNAs that bind them also might have changed little despite the enormous amount of evolution that has taken place all around them. (Winkler and Breaker 2007, 511)

10. That inanimate devices can be designed by agents to perform surrogate interpretations does not undermine the hypothesis, developed below, that the origin of interpretation was an important step in the emergence of life.

11. Localization is usually in some form of enclosed space such as a vesicle, although it may in the first instance be on a surface.

12. A recent report that the 5S RNA has a naturally occurring "leadzyme" activity may suggest a useful first line of inquiry (Barciszewska et al. 2003).

13. However, we do note the recent impressive work of Jack Szostak (2008) in obtaining replication of short sequences of DNA analogues using activated nucleotides trapped in vesicles.

14. In a subsequent article (Robinson and Southgate forthcoming b) we take the development of such a general theology of nature as a starting point and, using a semiotic framework, develop a specifically trinitarian approach to issues in the theology of creation, Christology, and anthropology.

REFERENCES

Barciszewska, M. Z., et al. 2003. "5S rRNA Is a Leadzyme. A Molecular Basis for Lead Toxicity." *Journal of Biochemistry* 133:309–15.

Deacon, Terrence W. 2006. "Reciprocal Linkage between Self-organizing Processes Is Sufficient for Self-reproduction and Evolvability." *Biological Theory* 1:136–49.

Deamer, David W. 1997. "The First Living Systems: A Bioenergetic Perspective." *Microbiology and Molecular Biology Reviews* (June): 237–61.

Gesteland, Raymond F., Thomas R. Cech, and John F. Atkins, eds. 2006. *The RNA World*. 3d ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Goodwin, B. 1995. *How the Leopard Changed Its Spots*. London: Orion.

Joyce, Gerald F. 2004. "Directed Evolution of Nucleic Acid Enzymes." *Annual Reviews in Biochemistry* 73:791–836.

Kauffman, Stuart A. 2000. *Investigations*. Oxford: Oxford Univ. Press.

Kauffman, Stuart, and Philip Clayton. 2006. "On Emergence, Agency, and Organization." *Biology and Philosophy* 21:500–520.

Küppers, Bernd-Olaf. 1990. *Information and the Origin of Life*. Cambridge: MIT Press.

Lui, L. T., Z. R. Yang, N. Lehman, B. Larson, A. J. N. Robinson, and C. C. B. Southgate. 2010. Environment-specific state-changes in simple protobiotic systems: Investigation of the role of "interpretative" responsiveness in the origin of life. Poster presented at the American Association for the Advancement of Science, San Diego, February. Available at *www.evolutioncreationsemiotics.org*.

Mandal, Maumita, and Ronald R. Breaker. 2004. "Gene Regulation by Riboswitches." Nature Reviews. *Molecular and Cell Biology* 5:451–63.

Orgel, Leslie E. 2002. "The Origin of Biological Information." In *Life's Origin: The Beginnings of Biological Information*, ed. J. W. Schopf, 140–57. Berkeley: Univ. of California Press.

Oyama, Susan, Paul E. Griffiths, and Russell D. Gray. 2001. "Introduction: What Is Developmental Systems Theory?" In *Cycles of Contingency: Developmental Systems and Evolution*, ed. S. Oyama, P. E. Griffiths, and R. D. Gray, 1-11. Cambridge: MIT Press.

Pohorille, A., et al. 2008. The Origins of Transmembrane Ion Channels. Address to the International Society for the Study of the Origin of Life Conference, Florence, 24–29 August.

Robinson, Andrew. In press. *God and the World of Signs: Trinity, Evolution and the Metaphysical Semiotics of C. S. Peirce*. Leiden: Brill.

Robinson, Andrew, and Christopher Southgate. 2010a. "Introduction: Toward a Metaphysic of Meaning." *Zygon: Journal of Religion and Science* 45:339–44.

———. 2010b. "A General Definition of Interpretation and Its Application to Origin of Life Research." *Biology and Philosophy* 25:163–82.

———. Forthcoming a. "Incarnation and Semiotics: A Theological and Anthropological Hypothesis. Part 1: Incarnation and Peirce's Taxonomy of Signs." *Theology and Science*.

———. Forthcoming b. "Incarnation and Semiotics: A Theological and Anthropological Hypothesis. Part 2: Semiotics, Anthropology, and Religious Transformation." *Theology and Science*.

———. Forthcoming c. "Semiotics as a Metaphysical Framework for Christian Theology." *Zygon: Journal of Religion and Science*.

Schneider, T. D. 2000. "Evolution of Biological Information." *Nucleic Acids Research* 28:2794–99.

Sebeok, Thomas. 1972. *Perspectives in Zoosemiotics*. The Hague: Mouton.

Short, T. L. 2002. "Darwin's Concept of Final Cause: Neither New nor Trivial." *Biology and Philosophy* 17:323–40.

———. 2004. "The Development of Peirce's Theory of Signs." In *The Cambridge Companion to Peirce*, ed. C. Misak, 214–40. Cambridge: Cambridge Univ. Press.

———. 2007. *Peirce's Theory of Signs*. Cambridge: Cambridge Univ. Press.

Szostak, J. 2008. What Can We Learn about the Origin of Life from Efforts to Design an Artificial Cell? Keynote address to the International Society for the Study of the Origin of Life Conference, Florence, 24–29 August.

Winkler, W. C., and R. R. Breaker. 2007. "Regulation of Bacterial Gene Expression by Riboswitches." *Annual Reviews in Microbio*logy 59:487–517.