Nova et Vetera, English Edition, Vol. 14, No. 2 (2016): 459–494

From Aristotle to John Searle and Back Again: Formal Causes, Teleology, and Computation in Nature

Edward Feser
Pasadena City College
Pasadena, CA

Introduction

Talk of information, algorithms, software, and other computational notions is commonplace in the work of contemporary philosophers, cognitive scientists, biologists, and physicists. These notions are regarded as essential to the description and explanation of physical, biological, and psychological phenomena. Yet, a powerful objection has been raised by John Searle, who argues that computational features are observer-relative, rather than intrinsic to natural processes. If Searle is right, then computation is not a natural kind, but rather a kind of human artifact, and is therefore unavailable for purposes of scientific explanation.

In this paper, I argue that Searle's objection has not been, and cannot be, successfully rebutted by his naturalist critics. I also argue, however, that computational descriptions do indeed track what Daniel Dennett calls "real patterns" in nature. The way to resolve this *aporia* is to see that the computational notions are essentially a recapitulation of the Aristotelian-Scholastic notions of formal and final causality, purportedly banished from modern science by the "mechanical philosophy" of Galileo, Descartes, Boyle, and Newton. Given this "mechanical" conception of nature, Searle's critique of computationalism is unanswerable. If there is truth in computational approaches, then this can be made sense of, and Searle's objection rebutted, but only if we return to a broadly Aristotelian-Scholastic philosophy of nature.

The plan of the paper is as follows. The next section ("From Scholasticism to Mechanism") provides a brief account of the relevant Aristotelian notions and of their purported supersession in the early modern period. The third section ("The Computational Paradigm") surveys the role computational notions play in contemporary philosophy, cognitive science, and natural science. The following section ("Searle's Critique") offers an exposition and qualified defense of Searle's objection to treating computation as an intrinsic feature of the physical world—an objection that, it should be noted at the outset, is independent of and more fundamental than his famous "Chinese Room" argument. In the fifth section ("Aristotle's Revenge"), I argue that the computational paradigm at issue essentially recapitulates certain key Aristotelian-Scholastic notions commonly assumed to have been long ago refuted and that a return to an Aristotelian philosophy of nature is the only way for the computationalist to rebut Searle's critique. Finally, in "Theological Implications," I explore ways in which computationalism, understood in Aristotelian terms, provides conceptual common ground between natural science, philosophy, and theology.

From Scholasticism to Mechanism

Scholastic thinkers, building on Aristotle, developed a complex network of interrelated concepts they regarded as essential to understanding the natural order. These include the distinctions between actuality and potentiality, substance and accidents, substantial form and prime matter, efficient and final causes, and so on. Contrary to a very common misconception, these notions are not in competition with scientific explanations as we now understand the notion of a scientific explanation. Rather, they are part of a metaphysical framework that, the Scholastic maintains, any possible scientific explanation must presuppose, and in light of which the results of scientific investigation must be interpreted.

I happen to think this framework is correct, and I have presented a thorough exposition and defense of it in my book *Scholastic Metaphysics*.[1] For present purposes, three Aristotelian-Scholastic notions are especially important: *substantial form*, *immanent teleology*, and *proportionate causality*. Let's consider each one in turn. A "substantial form"

---

1 Edward Feser, *Scholastic Metaphysics: A Contemporary Introduction* (Heusenstamm; Piscataway, NJ: Editiones Scholasticae; Transaction Publishers, 2014). For another recent book-length defense, see David S. Oderberg, *Real Essentialism* (London: Routledge, 2007).

is contrasted with an "accidental form," and the difference can be illustrated with a simple example. Consider a liana vine, which is the sort of vine Tarzan uses to swing around the jungle. Like any natural object, a liana vine has certain characteristic properties and operations. It takes in water and nutrients from the soil through its roots, exhibits distinctive growth patterns, and so forth. Suppose Tarzan makes a hammock using several living liana vines. The resulting object will also have certain distinctive features, such as being strong enough to support a grown man, being comfortable enough to take a nap in, and so on.

Now there is a clear difference between these two sets of features. The tendencies to take in water and nutrients, to exhibit certain growth patterns, and the like are *intrinsic to* or *"built into"* the liana vines. That is just what healthy vines will do when left to themselves. Functioning like a hammock, however, is not intrinsic to the vines, but *extrinsic* or *imposed from outside*. The vines not only have no "built in" tendency to function like a hammock, but over time will no doubt lose their suitability so to function unless Tarzan occasionally prunes and reties them. Taking in nutrients, growing in certain patterns, and the like, are *natural* to liana vines, but functioning as a hammock is not.

The Scholastic would express this difference by saying that to be a liana vine is to have a certain *substantial form*, whereas to be a hammock is to have only a certain *accidental form*. Artifacts are the stock examples of objects having only accidental forms, but it is important to emphasize that the distinction between accidental forms and substantial forms does not correspond exactly to the distinction between man-made objects and objects that exist "in the wild." Breeds of dog are in a sense man-made, but dogs still have substantial forms rather than accidental forms. A pile of stones that has formed randomly at the bottom of a hill is not man-made but still has only an accidental form rather than a substantial form, since the rocks have no intrinsic tendency to form a pile. It is not the fact of occurring "in the wild" rather than being man-made, but rather the having of an intrinsic rather than extrinsic tendency to manifest certain characteristic properties and activities that is the mark of the presence of a substantial rather than accidental form. A "natural" tendency, in the technical Scholastic sense, is one that follows from the having of a substantial rather than accidental form. Hence something man-made (such as children, new breeds of dog, or water synthesized in a lab) can be "natural" in the relevant sense, whereas some objects that occur "in the wild" (such as random piles of stones) are *not* "natural" in the relevant sense.

An intrinsic tendency of the sort in question confers a *unity* on the object that has it which makes of it a true *substance*. Objects having tendencies of only the extrinsic sort are, in the Scholastic view, not true substances, but rather accidental features of components which are themselves true substances. Hence a stone has a substantial form and is thus a true substance, but a pile of stones is not, the form of a pile being only an accidental form that several stones have taken on. A liana vine has a substantial form and is thus a true substance, whereas a hammock made out of liana vines is not, the form of a hammock being only an accidental form that several liana vines have taken on.

As Eleonore Stump has suggested, an indicator of the presence of a true substance, and thus of something having a substantial rather than merely accidental form, is the possession of *irreducible causal powers*.[2] A liana vine has causal powers that are irreducible to those of the cells that make it up, whereas the causal powers of a hammock are nothing over and above the aggregate of the causal powers of its component vines. Stone has causal powers that are irreducible to those of the molecules that make up the stone, whereas the causal powers of a pile of stones are nothing over and above the aggregate of the causal powers of the individual stones in the pile, and so forth. A difference of substantial forms thus entails a difference of substances, each with its distinctive causal powers.

There is an obvious way in which the notion of immanent teleology is implicit in what has been said so far. For something to exhibit *teleology* is for it to *point to* or be *directed toward* some end or outcome. For it to exhibit *immanent* teleology is for this directedness to be intrinsic to it, rather than imposed from outside. Liana vines are *directed toward* outcomes—drawing in nutrients and water, growing in certain patterns, and so forth—*intrinsically*, just by virtue of being the kind of thing they are. That is to say, they are directed toward those outcomes by virtue of having a certain substantial form. Liana vines are *directed toward* functioning as a hammock only *extrinsically*, by virtue of some end or outcome imposed on them from outside. That is to say, they are directed toward that end only by virtue of having a certain accidental form.

As this indicates, while writers like William Paley famously emphasize the similarity between organisms and artifacts, Aristotle and the

---

2 Eleonore Stump, "Emergence, Causal Powers, and Aristotelianism in Metaphysics," in *Powers and Capacities in Philosophy: The New Aristotelianism*, ed. Ruth Groff and John Greco (London: Routledge, 2013).

Scholastic writers who build on his work instead *contrast* organisms with artifacts. For the Aristotelian, organisms are *not* like artifacts precisely insofar as they have substantial forms and immanent teleology, while artifacts have only accidental forms and extrinsic teleology.[3] Paley and his followers also put a heavy emphasis on biological function, but for the Scholastic, biological function is just a special and more complex instance of a wider and usually much simpler phenomenon. Immanent teleology exists wherever there is causation. If a cause *A* regularly generates a certain specific effect or range of effects *B*, rather than *C* or *D*, or no effect at all, that can only be because generating *B* is the end or outcome toward which *A* intrinsically *points* or is *directed*. Hence, the phosphorus in the head of a match will, as long as the match has not been damaged, generate flame and heat when the match is struck rather than frost and cold, or rather than turning the match into a bouquet of flowers or a piece of spaghetti. Ice submerged in room-temperature water will cool it down rather than cause it to boil or to change it into oil, and so forth. Unless we suppose that the causes in question are inherently or of their nature *directed toward* those effects, then in the Scholastic view, we have no way of making intelligible why it is those specific effects that the causes in question in fact reliably produce.

In this way, what Aristotelians traditionally call *efficient* causality presupposes *final* causality. This Scholastic theme has been recapitulated in recent analytic philosophy among writers who argue for the reality of *causal powers* or *dispositions*. At least some such writers argue that dispositions are "directed toward" their manifestations. Hence, brittleness is *directed toward* shattering as its characteristic manifestation, solubility is *directed toward* dissolving as its characteristic manifestation, and so forth. Since this directedness is somewhat like the directedness of thoughts toward their objects, but in a way that involves no conscious awareness or mental content, and since it is taken by these theorists to be an intrinsic feature of the physical or natural world, George Molnar has labeled it "physical intentionality," and John Heil calls it "natural intentionality."[4] But it is essentially what the Scholastics had in mind in affirming final causality or teleology as immanent to the natural order.

---

[3] See Edward Feser, "Teleology: A Shopper's Guide," *Philosophia Christi* 12 (2010): 142–159.

[4] George Molnar, *Powers: A Study in Metaphysics* (Oxford: Oxford University Press, 2003), chapter 3; and John Heil, *From an Ontological Point of View* (Oxford: Clarendon Press, 2003), 221–222.

Now, if efficient causes point forward toward their characteristic effects, effects also point backward toward their causes. This brings us to what Scholastic writers sometimes call the *principle of proportionate causality*, according to which whatever is in an effect must in some way or another be in its total cause, whether "formally," "virtually," or "eminently" (to use the traditional jargon). Suppose you need twenty dollars and I give it to you. The effect is your having a twenty-dollar bill. It might be that the way I was able to bring this about is by virtue of actually having a twenty-dollar bill with me. That would be a case where what is in the effect was in the cause "formally"; that is to say, you have the "form" or fit the pattern of possessing a twenty-dollar bill because I actually had that very same form or fit that pattern myself. It could instead be, though, that I did not initially have twenty dollars on me, but I did have at least that much in my bank account and could go retrieve it on demand. In that case, what is in the effect was in the total cause (namely me and my bank account) "virtually." The twenty-dollar bill was not actually or formally present, but it could be generated at will. Or, to take a more exotic case, suppose I did not have even twenty dollars in the bank, but did have access to a U.S. Treasury printing press and ran off a brand new twenty-dollar note to give to you. In that case, although I did not have twenty dollars "formally" or "virtually," I did have it "eminently," in the sense that I had something even more fundamental, the power to generate a new twenty-dollar bill.

A lot more could be said, but that suffices to give us a sense of the relevant Scholastic ideas. Now, the early modern philosophers replaced this Aristotelian-Scholastic conception of nature with what is sometimes called a "mechanistic" conception. Much of what was originally associated with mechanism (such as the idea that all physical causation could be understood on a simple push-pull model) has been long abandoned, but the core idea that has persisted to the present day is that there is no irreducible teleology or final causality immanent to nature. A corollary is that there are no substantial forms (since, as we have seen, to have a substantial form entails having an *intrinsic tendency toward* the manifestation of certain characteristic properties and activities). The way this mechanistic alternative view was initially spelled out was in theological terms. Natural objects were reconceived by thinkers like Newton and Paley as divine artifacts. Just as Tarzan's hammock in our earlier example has only an accidental form and its function is an instance of extrinsic or externally imposed teleology, so too natural objects themselves were to be understood as machines on which a kind of accidental form and extrinsic teleology or function has been imposed by God. "Laws of nature" were, in effect, descriptions of the patterns of operation the divine artificer had put into his artifacts.

Later thinkers would, of course, remove God from the picture, but though that step is commonly treated as if it were a move away from obscurantism, in fact it only made the overall metaphysical situation murkier. For the Scholastics, there was a tight *intrinsic* metaphysical connection between the substances that make up the natural order and the events into which they enter. For, by virtue of a thing's substantial form, its causal powers *inherently* point toward certain effects as toward a final cause, and nothing in any effect could have gotten there unless it were in some way in its cause. The early, theologically-oriented "mechanical philosophers" replaced these intrinsic connections between objects and events with a set of *extrinsic* connections directly established by God—in effect, with divinely imposed accidental forms and extrinsic teleology. But if we abandon both the Aristotelian apparatus of immanent formal and final causes and the early modern conception of God as artificer, it seems we are left with neither an intrinsic nor an extrinsic source of the order in the world, and thus with no source of order at all.

That is, of course, exactly what we find in Hume, for whom all events are inherently "loose and separate." In Hume's view, and contrary to the Scholastic doctrines of substantial form and immanent teleology, any effect or none may follow upon any cause. Contrary to the Scholastic principle of proportionate causality, anything might, in principle, come into being with no cause whatsoever. Contrary to early modern "mechanical philosophers" like Descartes, Newton, and Paley, we cannot appeal to God either as the source of order in the world. Many contemporary philosophers would, of course, appeal to a non-theological version of "laws of nature" as an explanation, but that just raises the question of what a law of nature *is* if it is neither a shorthand description of the way a thing will operate given its substantial form (which is what the Scholastic would say) nor a direct divine decree (which is what Descartes, Newton, and company would say). The standard Humean view is that a "law of nature" is a statement of the regularities that happen as a matter of fact to exist in nature. The trouble is that, if that is what a law is, then it is really just a *re-description* in novel jargon of the order that exists in the world. And in that case, it cannot *explain* that order.

That the modern "mechanistic" conception of nature has left the status of the "laws of nature" so puzzling is, as we will see, arguably part of the reason for the appeal of the computational paradigm. Let us now turn to that.

The Computational Paradigm

Fundamental to the notion that natural processes are computational is the idea of *information*. The term "information" has become something of a buzzword in contemporary pop-science writing, and unfortunately it is not always used with precision. It is generally acknowledged, however, that the sense of the term operative in computer science, and thus in arguments to the effect that computational processes literally exist in nature, is not the everyday sense of the term, but rather a technical sense.

The technical sense in question is essentially the one associated with mathematician Claude Shannon's celebrated theory of information.[5] Shannon was concerned with information in a *syntactic* rather than *semantic* sense. Consider the *bit*, the basic unit of information, which has one of two possible values, usually represented as either 0 or 1. To consider a bit or string of bits (e.g., "11010001") in terms of some interpretation or meaning we have attributed to it would be to consider it semantically. Semantic information is the sort of thing we have in mind when we speak of "information" in the ordinary sense. To consider the properties a bit or string of bits has merely as an uninterpreted symbol or string of symbols is to consider it syntactically. This is "information" in the technical sense. When instantiated physically, a bit corresponds to one of two physical states, such as either of two positions of a switch, two distinct voltage levels, or what have you.

As David Chalmers points out, when physically instantiated, information in this technical sense essentially involves a causal correlation between a physical state of the sort in question and some effect at the end of a causal pathway leading from that state.[6] Think, for example, of the correlation between a switch's being either up or down and the light to which it is connected being either on or off. The position of the switch carries a single bit of information, and any physical

---

5 Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication* (Urbana, IL: University of Illinois Press, 1949).

6 David J. Chalmers, *The Conscious Mind* (Oxford: Oxford University Press, 1996), 281.

state that has the same effect down the causal pathway will carry the same information. Several switches (or, again, several distinct voltage levels or whatever) taken together will, naturally, carry more information. Following Chalmers, we can describe a combination of possible physical states (such as a combination of possible sets of positions of a number of switches) as an "information space." The structure of any information space will correspond to the structure of the set of possible effects down the causal pathway from the physical states that make up the information space.

Since it will be useful later on in our discussion, let me quote at length from Chalmers's example of the information carried by a compact disc. He writes:

> A disk [*sic*] has an infinite number of possible physical states, but when its effects on a compact-disk player are considered, it realizes only a finite number of possible information states. Many changes in the disk—a microscopic alteration below the level of resolution of the optical reading device, or a small scratch on the disk, or a large mark on the reverse side—make no difference to the functioning of the system. The only differences relevant to the disk's information state are those that are reflected in the output of the optical reading device. These are the differences in the presence of pits and lands on the disk, which correspond to what we think of as "bits." . . . The physical states of different pressings of the same recording will be associated with the same information state, if all goes well. Pressings of different recordings, or indeed imperfect pressings of the same recording, will be associated with different information states, due to their different effects. . . .
>
> Each "bit" on the compact disk has an independent effect on the compact disk player, so that each location on the disk can be seen to realize a two-state subspace of its own. Putting all these independent effects together, we find a combinatorial structure in the space of total effects of a compact disk, and so we can find the same combinatorial structure in the information space that the compact disk realizes.[7]

---

[7] Ibid., 282.

As this example indicates, the amount of "information," in the sense in question, that might be transmitted along a causal pathway is quantifiable, and that is what Shannon's theory of information is concerned with. That what comes out of the compact disc player when the disc is played counts as *music*—that the lyrics have a certain *meaning* and so forth—is completely irrelevant to how much information is transmitted. Again, "information" is being used here in a *syntactic* rather than *semantic* sense. What is at issue is what effects a bit or string of bits has considered merely as an uninterpreted symbol or string of symbols and entirely apart from what meaning or interpretation we assign to them.

Now, computers are said to process information. This is what happens when (to stick with Chalmers's compact disc example) you place a CD-ROM into your computer and the text file you have saved on it appears onscreen as a document written in English. Of course, you will not find anything that looks like English words on the CD-ROM. What happens is that the electrical states of the computer serve as a causal pathway by which the information state embodied in the CD-ROM generates the images on the screen. The information on the CD-ROM is the *input*, the images on the screen are the *output*, and the computer moves from the former to the latter because it is running an appropriate *algorithm* or set of instructions. But of course, you also will not find anything in the computer that looks like a set of instructions. The algorithm is *itself* embodied as information in the relevant sense—that is to say, as a certain configuration of electrical states. As biologist John Mayfield writes: "A computer can be seen as a device in which one state (the input) interacts with another state (the current machine configuration) to produce a final state (the output)."[8] *Computation* is just this transition from states that can be characterized as embodying an informational input via states that can be characterized as the embodiment of an algorithm into states that can be characterized as the output of the algorithm. A key property of computations is that you will not get more information out of them than went into them. As Mayfield puts it: "Algorithmic information shares with Shannon information the property that it cannot be created during a deterministic computation. The information content of the output can be less than that of the input, but not greater. Thus, algorithmic information conforms with our intuitive notion that information cannot be created out of thin air."[9]

---

8 John E. Mayfield, *The Engine of Complexity: Evolution as Computation* (New York: Columbia University Press, 2013), 45.

9 Ibid., 50.

Now, many contemporary philosophers and scientists hold that computation can be found not only in the machines we design for that purpose, but also in the natural world. In particular, the notions of *information*, *algorithms*, and the like have been claimed to have application to the understanding of phenomena studied in physics, biology, and neuroscience.[10] Consider physicist John Wheeler's famous "It from Bit" thesis.[11] The idea here is that, rather than physical states being metaphysically fundamental and information derivative, it is information (the "bit") that is metaphysically fundamental and the physical universe (the "it") that derives from information. Physicist Seth Lloyd and others have developed the theme into the suggestion that the universe just *is* a gigantic computer.[12] What exactly does all this mean, and why would anyone think it true?

Chalmers and physicist Paul Davies suggest illuminating interpretations. Following Bertrand Russell, Chalmers notes that physics does not tell us the intrinsic nature of the fundamental entities it posits: "Physics tells us nothing about what mass *is*, or what charge *is*: it simply tells us the range of different values that these features can take on, and it tells us their effects on other features."[13] Having mass or charge, like carrying syntactic information, is simply a matter of being in one of several states in a space of different possible states that might generate various outcomes at the end of causal pathways leading from those states. Now, if the fundamental entities of physics are essentially characterized in terms of their effects, and if to be information in the syntactical sense is just to have certain characteristic effects, then what physics gives us (Chalmers proposes) is essentially an informational conception of its fundamental entities.

Davies, noting that the idea of "laws of nature" is metaphysically problematic when removed from the theological context in terms of which Descartes and Newton understood it, proposes grounding laws instead in information considered as the "ontological basement"

---

[10] For a useful collection of papers on the subject, see *Information and the Nature of Reality: From Physics to Metaphysics*, ed. Paul Davies and Niels Henrik Gregersen (Cambridge, UK: Cambridge University Press, 2010).

[11] John Archibald Wheeler, with Kenneth Ford, *Geons, Black Holes, and Quantum Foam: A Life in Physics* (New York: W.W. Norton and Company, 1998), chapter 15.

[12] Seth Lloyd, *Programming the Universe* (New York: Vintage Books, 2006); see also Lloyd, "The Computational Universe," in Davies and Gregersen, *Information and the Nature of Reality*.

[13] Chalmers, *The Conscious Mind*, 302.

level of physical reality.[14] Physicist Rolf Landauer had put forward the thesis that the laws of physics are the algorithms according to which the universe computes.[15] Expounding Landauer's position, Davies notes that it opens up the possibility of seeing "the laws of physics [as] inherent in and emergent with the universe, not transcendent of it."[16]

When combined, Chalmers's and Davies's views suggest that the notion of the universe as a kind of computer provides a way of bringing the laws of physics "down to earth," as it were, and unifying them with the entities they govern. As we saw above, syntactic information is embodied in physical states correlated with some effect at the end of a causal pathway, and the algorithms by which this information is processed are themselves embodied as information and, thus, embodied in such physical states. If the universe is a kind of computer, then it is governed by the laws of nature in the same way a computer runs an algorithm, and the laws relate to the entities they govern the same way an algorithm is related to the physical states of a computer whose causal relations it describes. The puzzle about the status of the laws of nature that Hume left us with is thereby solved, or so it might seem.

The notions of information, algorithms, and the like have, if anything, played an even bigger role in biology. That genes carry syntactic or Shannon information about phenotypes is fairly uncontroversial, since this simply involves causal correlations between genetic factors and aspects of a phenotype. More controversial is whether there is *semantic* information to be found in biological phenomena—information with something comparable to the *meaning* or *intentional content* characteristic of thoughts and linguistic representations. Certainly biologists often describe the phenomena they study in ways that imply that there is such information. As philosopher of biology Alex Rosenberg notes:

> Molecular biology is . . . riddled with intentional expressions: we attribute properties such as being a *messenger* ("second messenger") or a *recognition* site; we ascribe *proofreading* and *editing* capabilities; and we say that enzymes can *discriminate* among substrates. . . . Even more tellingly . . . molecular developmental biology describes cells as having "positional information,"

---

14 Paul Davies, "Universe from bit," in Davies and Gregersen, *Information and the Nature of Reality*, 82.

15 Paul Davies, *The Mind of God* (New York: Simon and Schuster, 1992), 146–147.

16 Davies, "Universe from bit," 83.

> meaning that they *know* where they are relative to other cells and gradients. The naturalness of the intentional idiom in molecular biology presents a problem. All these expressions and ascriptions involve the representation, in one thing, of the way things are in another thing. . . . The naturalness of this idiom in molecular biology is so compelling that merely writing it off as a metaphor seems implausible. Be that as it may, when it comes to information in the genome, the claim manifestly cannot be merely metaphorical, not, at any rate, if the special role of the gene is to turn on its information content. But to have a real informational role, the genome must have intentional states.[17]

Now, whether intentionality or semantic content can be given a materialist explanation is itself controversial. Like other critics of materialism, I think it cannot be. Rosenberg also thinks it cannot be, but since he is a materialist, his solution is to take the eliminativist line, according to which intentionality and semantic content are illusions. Accordingly, he denies that there really is intentionality or semantic information to be found in biological phenomena. However, he also holds that "the crucial question is not intentionality but programming."[18] In particular, in Rosenberg's view, a genome still "programs the embryo" and runs "software," even if its doing so does not involve processing information of a semantic sort.[19] (Rosenberg happens to agree with the upshot of Searle's "Chinese Room" argument, according to which running a program is not sufficient for intentionality or semantics.)[20]

Others would go further. For example, philosopher of biology Peter Godfrey-Smith thinks that "genes 'code for' the amino acid sequence of protein molecules" in a sense that is appropriately regarded as semantic, though he adds that this "does not vindicate the idea that genes code for whole-organism phenotypes, let alone provide a basis for the wholesale use of informational or semantic language in biology."[21] Insofar as genes carry information vis-à-vis phenotypes,

---

17 Alex Rosenberg, *Darwinian Reductionism* (Chicago: University of Chicago Press, 2006), 99–100.

18 Ibid., 108.

19 Ibid., 107–108.

20 Ibid., 103–105.

21 Peter Godfrey-Smith, "Information in Biology," in *The Cambridge Companion to the Philosophy of Biology*, ed. David L. Hull and Michael Ruse (Cambridge, UK: Cambridge University Press, 2007), 109–110.

Godfrey-Smith thinks it only information of the syntactic or Shannon sort. Like Rosenberg, he also thinks there is at least a limited role for talk of "programs," in particular when describing the operation of gene regulation networks.[22]

Biologist Richard Dawkins is particularly eloquent on the subject of programs in nature. In *The Blind Watchmaker*, he writes:

> It is raining DNA outside. On the bank of the Oxford canal at the bottom of my garden is a large willow tree, and it is pumping downy seeds into the air. . . . The cotton wool is mostly made of cellulose, and it dwarfs the tiny capsule that contains the DNA, the genetic information. The DNA content must be a small proportion of the total, so why did I say that it was raining DNA rather than cellulose? The answer is that it is the DNA that matters . . . DNA whose coded characters spell out specific instructions for building willow trees. . . . It is raining instructions out there; it's raining programs; it's raining tree-growing, fluff spreading, algorithms. That is not a metaphor, it is the plain truth. It couldn't be any plainer if it were raining floppy discs.[23]

Others would go even farther in applying the notion of *semantic* information within biology.[24] For present purposes, however, we can simply note that at least the core computational notions of *syntactic* information and algorithms are widely applied within biology. Natural selection itself has been claimed by philosopher Daniel Dennett and biologist John Mayfield to amount to a kind of algorithm, and the evolutionary process to constitute a kind of computation or information processing.[25] The ubiquity within biology of the syntactic notion of information suffices for the purposes of the present paper, and what I will have to say about syntactic information will, if correct, be even more obviously true of semantic information.

But before moving on to a critical evaluation of the use of computational notions in physics and biology, let us briefly note one further area in which they are thought to have application, namely, in the

---

22 Ibid., 111–112.

23 Richard Dawkins, *The Blind Watchmaker* (New York: W.W. Norton and Company, 1987), 111.

24 See, e.g., the essays by John Maynard Smith, Terrence Deacon, Bernd-Olaf Küppers, and Jesper Hoffmeyer in Davies and Gregersen, *Information and the Nature of Reality*.

25 Daniel C. Dennett, *Darwin's Dangerous Idea* (New York: Simon and Schuster, 1995), and Mayfield, *The Engine of Complexity*.

study of the mind. The best-known instance of this approach is the idea that the mind is a kind of software and the brain a kind of computer hardware that runs the software. The former thesis, that the mind is a kind of software, is one that Searle labels "strong artificial intelligence" (or "strong AI"), and it also goes by the name "Turing machine functionalism." The latter thesis, that the brain is a kind of digital computer, is one that Searle labels "cognitivism." Obviously the theses are related, but they are distinct, and Searle has presented distinct arguments against each.

His famous "Chinese Room" argument is directed against the first, the "strong AI" thesis. I will have nothing to say about that argument here because I think Searle is simply and without qualification correct to hold that the mind is not a kind of computer program or software, though my reasons go beyond (even if they include) the ones he gives in that argument.[26] For present purposes, I want to focus instead on the "cognitivist" claim that the brain is a kind of computer, such that computation must be at least part of the story in a scientific account of human cognition, even if it is not the whole story.

In the work of philosophers like Paul Churchland, you will often find claims like the following:

> The brain represents the world by means of very high-dimensional *activation vectors*, that is, by a pattern of activation levels across a very large population of neurons. And the brain performs computations on those representations by effecting various complex *vector-to-vector transformations* from one neural population to another. This happens when an activation vector from one neural population is projected through a large matrix of synaptic connections to produce a new activation vector across a second population of nonlinear neurons.[27]

This approach to studying the brain is developed in great detail in works of computational neuroscience.[28]

---

[26] Edward Feser, "Kripke, Ross, and the Immaterial Aspects of Thought," *American Catholic Philosophical Quarterly* 87 (2013): 1–32.

[27] Paul M. Churchland, "Activation Vectors vs. Propositional Attitudes: How the *Brain* Represents Reality," in Paul M. Churchland and Patricia S. Churchland, *On the Contrary: Critical Essays, 1987–1997* (Cambridge, MA: The MIT Press, 1998), 41.

[28] See, e.g., Patricia S. Churchland and Terrence J. Sejnowski, *The Computational Brain* (Cambridge, MA: The MIT Press, 1992).

Now, if by "representations" such writers had in mind something like thoughts with conceptual content, then I think these sorts of claims would be false. In my view, the conceptual content of our thoughts cannot be explained in causal terms, or in any other terms acceptable to the materialist.[29] However, if what is meant is merely that there is information in the brain of the syntactic, Shannon sort, then the computationalist approach is certainly no less plausible here than it is in the case of physics or biology. Indeed, there can hardly be any doubt that the neural properties and processes described in such detail in books of computational neuroscience are real and important.

But is the specifically *computationalist* conceptual apparatus, here or in the other contexts we have considered, necessary to a correct description of the phenomena? Or is it just a dispensable and, indeed, misleading set of metaphors? This brings us at last to Searle's argument against cognitivism.

Searle's Critique

Again, the argument in question is not to be confused with Searle's famous "Chinese Room" argument.[30] In that argument, Searle's claim was that running a program does not entail having intentional content or meaning; as he famously summed it up: "syntax is not sufficient for semantics." Even if the brain could be said to process information in the syntactic sense Shannon was interested in, the "Chinese Room" argument entails that that would never by itself amount to the having of semantic information of the sort characteristic of thought. But the argument leaves open the question of whether the brain really does process information in at least the syntactic sense.

Searle's later argument against what he calls cognitivism is intended to show that it does not.[31] It aims to show that computation is not only not the *whole* story about what the brain does; it is not even *part* of the story. The basic idea of the argument is very simple. Whatever

---

29 Edward Feser, "Hayek, Popper, and the Causal Theory of the Mind," in *Hayek in Mind: Hayek's Philosophical Psychology*, ed. Leslie Marsh (Bingley: Emerald Group Publishing Limited, 2011); see also Feser, "Kripke, Ross, and the Immaterial Aspects of Thought."

30 John R. Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3 (1980): 417–424; see also Searle, *Minds, Brains, and Science* (Cambridge, MA: Harvard University Press, 1984).

31 John R. Searle, *The Rediscovery of the Mind* (Cambridge, MA: The MIT Press, 1992), chapter 9; see also Searle, "Is the Brain a Digital Computer?" in *Philosophy in a New Century: Selected Essays* (Cambridge, UK: Cambridge University Press, 2008).

else computation in the sense we are discussing might involve, at the very least it involves the physical instantiation of *symbols* or strings of symbols, whether 0s and 1s or some other kind of symbol. If left uninterpreted, the symbols will not carry *semantic* information. They will still constitute *syntactic* information, but only insofar as we do think of them as *symbols*, even if uninterpreted ones. The syntactical rules that make up the algorithm according to which the inputted symbols generate a certain output are rules that govern physical states precisely *qua symbols*. For example, they will be rules according to which the computer will give a 0 as output when it gets a 1 as input, or whatever. And that the computer instantiates a certain algorithm will, as we have seen, itself amount to there being certain further physical states that count as instances of certain symbols or bits of syntactic information. So, computation boils down to the instantiation of symbols.

The problem is this. The status of being a "symbol," Searle argues, is simply not an objective or intrinsic feature of the physical world. It is purely conventional or observer-relative. And thus the status of being something that is running an "algorithm," or "processing information," or "computing," is also conventional or observer-relative, rather than an intrinsic and objective feature of any physical system. This is obviously true where the computers of everyday experience are concerned. What they do constitutes the "processing" of "symbols," or "bits" of "information," according to an "algorithm" only because human designers and users of the machine *count* the electrical states as symbols, the transitions between states as the implementation of an algorithm, and so on. But the same thing is true of anything else we might think of as a computer—a brain, a genome, or the universe as a whole. Its status as a "computer" would be observer-relative because a computer is simply not a "natural kind," but rather a sort of artifact. Searle draws an analogy:

> We might discover in nature objects which had the same sort of shape as chairs and which could therefore be used as chairs; but we could not discover objects in nature which were functioning as chairs, except relative to some agents who regarded them or used them as chairs.

Similarly, he says:

> We could no doubt discover a pattern of events in my brain that was isomorphic to the implementation of the vi program on this

> computer. But to say that something is *functioning as* a computational process is to say something more than that a pattern of physical events is occurring. It requires the assignment of a computational interpretation by some agent.[32]

So, if a brain is a kind of computer, then that can, in Searle's view, be true only in the trivial sense that we can *interpret* various brain states as symbols and various neural processes as computations if we like. But in that sense all sorts of other things are "computers" too. Searle writes:

> For any program there is some sufficiently complex object such that there is some description of the object under which it is implementing the program. Thus for example the wall behind my back is right now implementing the Wordstar program, because there is some pattern of molecule movements which is isomorphic with the formal structure of Wordstar. But if the wall is implementing Wordstar then if it is a big enough wall it is implementing any program, including any program implemented in the brain.[33]

But if the brain or any other natural system (such as the genome or the universe as a whole) is computing only in the trivial and uninteresting sense in which a wall is "computing," then it is not computing in any sense that might be explanatorily useful in science or philosophy.

In short, Searle says, "computational states are not *discovered within* the physics, they are *assigned to* the physics."[34] They are no more a part of the furniture of the natural order of things than chairs are. Hence, just as no physicist, biologist, or neuroscientist would dream of making use of the concept of a chair in explaining the natural phenomena with which they deal, neither should they make use of the notion of computation.

Now, an objection frequently raised against Searle is that more is required of something if it is to count as a computer than merely that we could interpret some isolated set of its states as a computation. It also has to have the right kind of causal organization.[35] It is not

---

32 Searle, "Is the Brain a Digital Computer?" 95.
33 Ibid., 93.
34 Searle, *The Rediscovery of the Mind*, 210.
35 Ned Block, "Searle's Arguments against Cognitive Science," in *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, ed. John Preston and Mark Bishop (Oxford: Clarendon Press, 2002), 76–78; Chalm-

enough, for example, for a system plausibly to count as implementing the computation "1 + 2 = 3" that it has states corresponding to "1" and "+" and "2" and "=" that are followed by a state corresponding to "3." For what it does genuinely to count as addition, it must also be true that, had the input been states corresponding to "2" and "+" and "3" and "=," the output would have been a state corresponding to "5," that had the input been states corresponding to "3" and "-" and "2" and "=," the output would have been a state corresponding to "1," and so on for other counterfactual inputs and outputs. We need an isomorphism not just between this or that particular computation and this or that particular state of the system, but between, on the one hand, the structure of a program as a whole, and on the other, the causal structure of the entire physical system over time. And this will rule out cases like Searle's example of his wall implementing Wordstar.

But this, it seems to me, is not a serious objection to Searle. Searle acknowledges that a system's having an appropriate causal structure is a *necessary* condition for its implementing a program.[36] His point is that it is not a *sufficient* condition. To recall his parallel example, having an appropriate causal structure is also a necessary condition for something's being a chair. Wood and steel have such a structure, but shaving cream, cigarette smoke, and liquid water do not, since they lack the solidity and stability to hold someone up. But whether some wooden or steel object counts as a chair is still observer-relative, a matter of convention. Similarly, even though a system has to have the requisite causal structure in order to count as a computer, Searle's point is that it still will not count as one unless some observer assigns a syntactical interpretation to its physical states.[37]

That a physical system's having the appropriate causal structure can only ever be a necessary, and not a sufficient, condition for its implementing a program is given further support by an anti-computationalist argument from Saul Kripke that is related to but distinct from

---

ers, *The Conscious Mind*, 219–220; Ronald P. Endicott, "Searle, Syntax, and Observer Relativity," *Canadian Journal of Philosophy* 26 (1996):101–122, at 103–107; Josef Moural, "The Chinese Room Argument," in *John Searle*, ed. Barry Smith (Cambridge: Cambridge University Press, 2003), 234–235; and Georges Rey, "Searle's Misunderstandings of Functionalism and Strong AI," in Preston and Bishop, *Views into the Chinese Room*, 215–217.

36 Searle, *The Rediscovery of the Mind*, 209.

37 Edward Feser, *Philosophy of Mind: A Beginner's Guide* (Oxford: Oneworld Publications, 2006), 163.

Searle's.[38] Consider Kripke's "quus" function, which he defines as follows:

x quus y = x + y, if x, y < 57; = 5 otherwise.

The primary use Kripke makes of this odd example is, of course, to generate his famous skeptical paradox about meaning. A person's linguistic utterances and other behavior and the words and images he calls before his mind might all seem to show that he is adding when he says "1 + 2 = 3" and the like. But Kripke imagines a bizarre skeptic suggesting that for all we know, the person might really be carrying out "quaddition" rather than addition. If the person has never computed numbers higher than 57, then although we expect that, when he computes "68 + 57," his answer will be "125," it may be that he is quadding rather than adding, so that the answer will actually be "5." Nor would it matter if he *had* computed numbers higher than 57. For there is always some number, even if an extremely large one, equal to or higher than which he has never calculated, and the skeptic can always run the argument using that number instead. Nor would it matter if the person in question said, "I am adding and not quadding!"—because just as *we* might be misinterpreting his use of words like "plus" and "adding," so too might he be misinterpreting his *own* use of those terms.

Now, one reason Kripke's paradox is philosophically interesting is that it might be claimed to show that there is no fact of the matter about what we mean by our utterances. I do not think it really does show this, for reasons I have explained elsewhere.[39] But Kripke thinks it also has application as an argument against computationalism, and this seems to me correct. For, whatever we say about what *we* mean when we use terms like "plus," "addition," and so on, there are no physical features of a *computer* that can determine whether *it* is carrying out addition or quaddition, no matter how far we extend its outputs. No matter what the past behavior of a machine has been, we can always suppose that its next output—"5," say, when calculating numbers larger than any it has calculated before—might show that

38 Saul A. Kripke, *Wittgenstein on Rules and Private Language* (Cambridge, MA: Harvard University Press, 1982), 35–37; Jeff Buechner, "Not Even Computing Machines Can Follow Rules: Kripke's Critique of Functionalism," in *Saul Kripke*, ed. Alan Berger (Cambridge, UK: Cambridge University Press, 2011).

39 Feser, "Kripke, Ross, and the Immaterial Aspects of Thought."

it is carrying out something like quaddition rather than addition. Of course, it might be said in response that, if this happens, that would just show that the machine was malfunctioning rather than performing quaddition. But Kripke points out that whether some output counts as a malfunction itself depends on what program the machine is running, and whether the machine is running the program for addition rather than quaddition is precisely what is in question.

Obviously Kripke's argument raises questions of its own.[40] It suffices for present purposes to note how it bolsters Searle's point. Even if a physical system's having a certain causal organization is a necessary condition for its implementing a program for addition, it cannot be a sufficient condition because that causal organization will also be consistent with its implementing quaddition rather than addition. And the point is completely general. There will be parallel quaddition-like counterexamples for *any* claim to the effect that a physical system's having a certain causal structure is sufficient for its implementing some specific program.

An objection raised against Searle by John Haugeland is that the claim that syntactical features are observer-relative is falsified by the fact that there are empirical tests based on stringent specifications for whether something possesses such features.[41] A related objection is raised by Jeff Coulter and Wes Sharrock when they write that it is odd for Searle to suggest that "computation is merely an 'observer-relative' feature of a computer."[42] The point of these objections seems to be that, given how syntax and computation are defined, whether or not something has syntactical features or is a computer is just a straightforward factual matter. But this just misses Searle's point. He is not denying that there might be rigorously specifiable empirical criteria for whether something has syntactical features or is a computer. He is saying that, whether there are or not, that something fitting those criteria counts as a computer is ultimately a matter of convention, rather than observer-independent facts. Even if we had rigorously articulated empirical criteria for whether something is a chair, that anything counts as a chair in the first place would still be a matter of

---

40 For a detailed exposition and defense, see Buechner, "Not Even Computing Machines Can Follow Rules."

41 John Haugeland, "Syntax, Semantics, Physics," in Preston and Bishop, *Views into the Chinese Room*, 391.

42 Jeff Coulter and Wes Sharrock, "The Hinterland of the Chinese Room," in Preston and Bishop, *Views into the Chinese Room*, 196.

convention, and thus observer-relative. Searle is saying that the same thing is true of whether something possesses syntactical or computational features.[43]

Ronald Endicott objects that the claim that the symbols posited by the computationalist are observer-relative rests on a false analogy with the symbols of natural languages.[44] True, the symbols of English, German, and the like have the meanings they do only because they are assigned meanings as a matter of convention. But the symbols posited by the computationalist, says Endicott, are not like that. The computationalist takes them to get their meaning, instead, in a way described by some naturalistic theory of meaning, such as a causal covariation theory. And the biologist's talk of DNA codes and the like shows, in Endicott's view, that it is possible for there to be symbols in nature apart from interpreters.

But this line of objection both begs the question and misses the point. It begs the question in two ways. First, Searle has argued that causal covariation and other naturalistic theories of meaning all fail, so a critic can hardly take such a theory for granted when criticizing Searle.[45] Second, whether DNA and related biological phenomena *literally* can be said to have computational features is, at least implicitly, precisely part of what is at issue between Searle and his critics. That computational notions are useful to the biologist would presumably be regarded by Searle as comparable to the fact that there are naturally occurring objects that we find it comfortable to sit on. The latter fact does not entail that it is not a matter of convention whether something is a chair, and the former fact (Searle would presumably say) does not entail that it is not a matter of convention whether DNA, the brain, or anything else counts as a computer.

Endicott misses the point when he speaks as if the issue has to do with whether the symbols posited by the computationalist get their meaning in something like the way the symbols of natural languages do, or instead in the way naturalistic theories of meaning say they do. Semantics was the topic of Searle's Chinese Room argument, but not

---

43 See Jeff Buechner, *Gödel, Putnam, and Functionalism: A New Reading of Representation and Reality* (Cambridge, MA: The MIT Press, 2008), 158–159.

44 Endicott, "Searle, Syntax, and Observer Relativity," 111.

45 Searle, *The Rediscovery of the Mind*, 49–52. I think Searle is correct to reject such theories, for reasons set out in Feser, "Hayek, Popper, and the Causal Theory of the Mind," and Feser, "Kripke, Ross, and the Immaterial Aspects of Thought."

of his argument against what he calls cognitivism. Here the issue is not how the symbols posited by the computationalist get their meaning, but rather whether it even makes sense in the first place to speak of symbols—even *uninterpreted* symbols—existing apart from human convention and apart from any observer.

Jeff Buechner raises several further objections against Searle.[46] First of all, Buechner notes that computations, like the objects of mathematical discourse, are abstract objects. Now, since conventionalist theories of mathematics are highly problematic, it is no less problematic to treat computation as if it were merely conventional or observer-relative.[47] It seems to me that the obvious retort to this is that it just misunderstands what Searle is saying. Searle is not saying that computations considered as abstract objects are conventional or observer-relative; he is saying that whether such-and-such a physical system *implements* a computation is conventional or observer-relative. Buechner considers this possible reply.[48] He concedes that, in a physical system built by an engineer, there is a sense in which the fact that it implements a computation is observer-relative. But he suggests that such a system could instead be "assembled through evolutionary pressures" and implies that, if Searle were to insist that an intelligent designer would be necessary for such a system to count as implementing a computation, he would be committing himself to an untestable "Intelligent Design" theory.

But the problem with Buechner's response can be seen by once again considering Searle's parallel example of a chair. Something that just happened to be the sort of thing we would find it comfortable to sit on could, in principle, come about by evolutionary processes. All the same, it would not count as a *chair* unless some observer decided so to count it, for chairs are not natural kinds, but products of convention. Similarly, Searle need in no way deny that something as complicated as the computers that human engineers construct could come about via evolutionary processes. He would deny only that this would, apart from an observer who assigns a computational interpretation to it, count as a computer. This no more commits him to "Intelligent Design" theory than does denying that a chair-like object that arose via evolutionary processes would, in the strict sense, really be a chair commit one to "Intelligent Design" theory. And if Buechner

---

[46] Buechner, *Gödel, Putnam, and Functionalism*, chapter 5.
[47] Ibid., 160–165.
[48] Ibid., 166–168 and 324n.11.

digs in his heels and insists that such a product of evolution *must really be* a computer and not merely something to which we might assign a computation interpretation—that is to say, if he says that it meets conditions *sufficient*, and not merely *necessary*, for being a computer—then he is just begging the question against Searle.

Buechner also concedes that there is a degree of convention or observer-relativity in symbols and syntax, just as there is in the system of numerals we use to do mathematics, but he thinks this does not suffice to establish Searle's position.[49] With any system of numerals, "the laws of arithmetic must be respected and human limitations must be respected."[50] What Buechner has in mind by the first constraint, it seems, is that if a system of numerals allowed us to count as true a statement like "2 + 2 = 5" (for example), then it would obviously be deficient for the purposes of doing arithmetic. What he has in mind by the second constraint is that certain symbols would not be useful for us for the purposes of doing mathematics given, for example, our perceptual limitations. For instance, we just find it harder to read "||||||||" than "8" (and so on for other numbers), and thus it would be practically impossible for us to do arithmetic via a stroke notation instead of a decimal notation. Now, since it is not within our power to change these constraints, the choice of which numerals to use is not entirely observer-relative, even if it is to some degree. But the same point can be made about the choice of syntax and symbols.

The problem with this as a reply to Searle is that, once again, all that has been shown is something Searle has already conceded, which is that having a certain physical structure is a *necessary* condition for a system's carrying out a certain computation. Searle's point, though, is that it is nevertheless not a *sufficient* condition. To be sure, Buechner adds a further point that, whereas in the case of computers made by us, human designers choose which symbols to use within the constraints in question, in the case of a natural computer like a brain, it is evolution that "chooses" among the various possible symbols fulfilling the constraints in question. He adds that these symbols could still be there even if we do not recognize them as such. Once again, though, Buechner's appeal to evolution either is a *non sequitur* or begs the question. If evolution produced something that was chair-like, it would not follow that it had produced a chair, and if evolution produced something symbol-like, it would not follow that it had produced

---

49 Ibid., 169–172.

50 Ibid., 170.

symbols. And if Buechner simply insists that this would follow, then he is assuming precisely what is at issue, since Searle's whole point is that no natural processes—including evolution—could of themselves produce something that was literally a computer.

Aristotle's Revenge

So, Searle's critics have failed to rebut his argument against cognitivism successfully. That is not to say, however, that they have not a leg to stand on. For Searle's critique to be decisive, he needs not only to give an argument *against* the claim that computation is intrinsic to the natural world, but also to show that there are no good positive arguments *for* the claim that it is intrinsic to the natural world. Are there any good positive arguments for that claim?

It seems to me that there are. We have already seen why at least some physicists, biologists, and neuroscientists would characterize the phenomena they study in terms of notions like *information*, *algorithms*, and the like. Searle would have to say that these are at best merely useful fictions and that everything that has been put in these computational terms could be said without recourse to them. But that does not seem to be the case. To see why, first consider once again Kripke's "quus" example. Kripke's skeptic claims that there is no fact of the matter about whether any of us is ever doing addition rather than "quaddition." The common-sense view, of course, is that there is a fact of the matter and that the fact is that we are doing addition and not "quaddition." Hence, if we came across someone whose arithmetical behavior seemed perfectly normal except that, when he calculated "68 + 57," his answer was "5" instead of "125," we would not conclude that he really was "quadding" after all. We would conclude, instead, that he was doing addition but, in this case, doing it badly. We would regard his answer as the result of a typographical error or momentary confusion, or perhaps even delirium, temporary insanity, dementia, brain damage, or what have you. We would not think of him as a properly functioning system carrying out "quaddition," but as a malfunctioning system carrying out addition.

We need not worry for present purposes about what would rationally *justify* our taking this view, since meaning skepticism is not our topic.[51] What I want to consider here is the following sort of parallel case. Recall some of the computationalist claims I cited earlier.

---

[51] Again, see Feser, "Kripke, Ross, and the Immaterial Aspects of Thought," for further discussion of Kripke's paradox.

Rosenberg says that the genome "programs the embryo." Churchland says that "the brain represents the world by means of very high-dimensional activation vectors" and "performs computations on those representations." Now, for the program Rosenberg says the genome is running, for the program Churchland says the brain is running, and for any other program someone wants to attribute to a natural process, we can construct a quaddition-like paradox. For instance, we can imagine what we might call a "quembryo" program that, when the genome runs it, produces the same results that the embryo program does except that the embryo does not develop eyes. Now, consider a human embryo that never develops eyes. Should we say that the genome that built this embryo was running what Rosenberg would call the embryo program but that there was a malfunction in the system? Or should we say instead that the genome was actually running the "quembryo" program and that there was no malfunction at all and things were going perfectly smoothly?

A Kripke-like skeptic would, of course, say that there is no fact of the matter. But notice that Searle, since he holds that there are no programs at all really running here in the first place, would also have to say that there is no fact of the matter. But that simply does not seem plausible. If there really is such a thing as a difference between a properly functioning organism and a malfunctioning one, then it seems to follow that Rosenberg's postulated embryo program captures something about the facts of the situation that our imagined "quembryo" program does not. Using computer jargon, our hypothetical "quembryo" skeptic might say of the embryo's lack of eyes: "Maybe that's not a bug, but a feature!" But he would be wrong. The lack of eyes *is* a bug, and not a feature. Unless we are skeptics about the very distinction between properly functioning and malfunctioning organisms—and it is hard to see how biology would be possible given such skepticism—then it seems we have to agree that there really is something to the claim that what we have here is a malfunctioning system running the embryo program, as opposed to a properly functioning system running the "quembryo" program.

If the computer scientist's distinction between "bugs" and "features" has application to natural phenomena, so too does the distinction between "software" and "hardware." For (to stick with the embryo example) the lack of eyes is as dysfunctional in one human embryo as it is in another. A natural way of putting this is that all human embryos are running the *same* program, that the very same software is, as it were, being implemented in different pieces of hardware. That is why what is a "bug" or "feature" for one embryo is also a "bug" or "feature" for the others. Searle's view seems to be that there is nothing true in the computationalist's description of a natural physical system that cannot be captured by a description of the causal processes taking place in the system. But that does not seem to be correct. For there is a distinction to be made between *normal* and *aberrant* causal processes, and there is a distinction to be made between a *general type* of normative causal process and specific *token instances* of that type. The computationalist's language captures these distinctions in a way that a mere description of which causal processes happen to be taking place does not.

To borrow some jargon from Daniel Dennett, Searle supposes, in effect, that everything that is true of a natural object or process can be captured by taking the "physical stance" toward it, but in fact there are aspects that can be captured only by taking the "design stance," and these are precisely the ones captured by the computationalist description.[52] The computationalist description captures what Dennett calls "real patterns" in nature, patterns irreducible to the purely causal description to which the "physical stance" confines itself.[53] For it is only by taking the "design stance," which is defined by consideration of *proper function*—where regarding the genome as running embryo software rather than "quembryo" software is at least one way of doing this—that we can make sense of the facts that the lack of eyes is an aberration and that this is true of human embryos as such, not just of this or that particular embryo.

Now, if all of this is correct, then we seem to have what Aristotle calls an *aporia*, a puzzle arising from the existence of apparently equally strong arguments for two or more inconsistent claims—in this case, equally strong arguments both against and for the claim that there is computation in nature. And the way to resolve it, I suggest, is to see that, while Searle's position is unavoidable if we take for granted the essentially "mechanistic" conception of nature to which he and his naturalist critics are both committed, the computationalist approach can be made sense of if we adopt instead a broadly Aristotelian-Scholastic conception of nature. (Or perhaps it would be more precise to say that, given an Aristotelian-Scholastic metaphysics, we can make

---

52 Daniel C. Dennett, "Intentional Systems," in *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge, MA: The MIT Press, 1981).

53 Daniel C. Dennett, "Real Patterns," in *Brainchildren: Essays on Designing Minds* (Cambridge, MA: The MIT Press, 1998).

sense of the idea that there is, in nature, something *analogous* to computation, in the technical Thomistic sense of analogy.)[54] In their use of computational notions, contemporary naturalists have unwittingly recapitulated the formal and final causality that they, like their early modern "mechanical philosophy" forebears, thought had been banished for good.

Recall that "information" in the technical, syntactic sense essentially involves a causal correlation between a physical state and some effect at the end of a causal pathway leading from that state. Now, any physical state has any number of effects along a causal pathway. For instance, the physical states of the compact disc in the example from Chalmers cited earlier have, among their effects, the sounds that come out of the CD player when the disc is played. But those physical states have many other effects as well. For example, there is the electrical activity that occurs in the circuitry of the CD player, which in turn causes the sounds to emerge from the speakers, and there is the shaking of the nearby walls that might take place if the volume is turned up too loud. Now, we say that the physical states of the compact disc carry "information" specifically about the *sounds* they cause, rather than about the electrical activity in the CD player or the shaking of the walls. The reason for that, of course, is that the designers of compact discs made them for the *purpose* of allowing us to play back the sounds in question, rather than for the purpose of generating electrical activity or causing walls to shake. It is the existence of that purpose that allows us to identify the sounds as the specific effect down the causal pathway about which the physical states of the compact disc carry information.

But no such observer-relative purposes can be appealed to in the case of the information the computationalist attributes to physical states occurring in nature. That is, of course, why Searle says there is no information to be found in such states. But if we suppose that Aristotelian teleology is a real feature of nature, after all, then we can make sense of such naturally occurring information. In particular, if we suppose that a physical state of type *S* inherently "points to" or is "directed at" some particular type of effect *E* down the causal pathway—rather than to some earlier effect *D* or some later effect *F*—then

---

[54] I owe the suggestion that computational descriptions be understood as analogical to Steven W. Horst, *Symbols, Computation, and Intentionality: A Critique of the Computational Theory of Mind* (Berkeley and Los Angeles: University of California Press, 1996), 6–7. For an overview of the Thomistic doctrine of analogy, see Feser, *Scholastic Metaphysics*, 256–263.

we have a way of making intelligible how *S* carries information about *E* rather than about *D* or *F*. Without such teleology, though, it is hard to see why there would be anything special about *E* by virtue of which it would be the effect about which *S* carries information.[55]

Consider also that, when we speak of a pocket calculator running a program or algorithm for addition rather than "quaddition," it is easy to make sense of this, given that the designers of the calculator designed it for the purpose of doing addition rather than quaddition. But how do we make sense of the genome running the embryo program rather than the "quembryo" program, given that there is no human observer who assigns this purpose to it? If we suppose that there are such things as Aristotelian substantial forms after all, then we have a way of making this intelligible. For there to be a fact of the matter that the genome is running the embryo program rather than the "quembryo" program is for the genome to have the sort of *intrinsic* tendency toward certain characteristic operations that distinguishes a substantial form from a merely accidental form.[56] Moreover, as John Mayfield notes, "an important requirement for an algorithm is that it must have an outcome,"[57] and "instructions" of the sort represented by an algorithm are "goal-oriented."[58] The genome algorithm, for example, has as its "goals" or "outcomes" the sorts that are characteristic of the embryo program, rather than the "quembryo" program. It is hard to make sense of this except as an instance of Aristotelian immanent teleology.

Other aspects of the computationalist conception of nature also echo the Aristotelian-Scholastic conception. For instance, when Mayfield notes (as we saw earlier that he does) that "the information content of the output [of a computation] can be less than that of the input, but not greater," he is essentially recapitulating the principle of proportionate causality, according to which whatever is in an effect must in some way or other be in its total cause.

Some of the moves made by Searle's critics also at least gesture, inadvertently, in a broadly Aristotelian direction. For example, in

---

55 See Feser, "Hayek, Popper, and the Causal Theory of the Mind."

56 See James F. Ross, "The Fate of the Analysts: Aristotle's Revenge," *Proceedings of the American Catholic Philosophical Association* 64 (1990): 51–74; see also James Ross, *Thought and World: The Hidden Necessities* (Notre Dame, IN: University of Notre Dame Press, 2008), chapter 7.

57 Mayfield, *The Engine of Complexity*, 44.

58 Ibid., 13.

response to Searle's claim that his wall is running Wordstar, Endicott objects that, before we can plausibly attribute a program to some physical system, we have to consider "*non-gerrymandered* physical units" and "a *physical system whose parts have the disposition to causally interact in the way specified by the program*."[59] But a "non-gerrymandered" physical unit occurring in nature arguably suggests one marked-off from others by virtue of having a substantial form, and a part having a "disposition" causally to act in certain specific ways arguably suggests one that is "directed toward" a certain kind of manifestation as toward a final cause.

Other writers have explicitly noted the Aristotelian implications of computational descriptions of natural phenomena. The neuroscientist Valentino Braitenberg has said that "the concept of information . . . is Aristotle *redivivus*, the concept of matter and form united in every object of this world."[60] Philosopher of science John Wilkins calls information "the new Aristotelianism" and the "New Hylomorphism," though unlike Braitenberg, he does so disapprovingly, considering the notions in question to entail a regress to an outmoded conception of nature.[61]

I imagine that Searle would share Wilkins's attitude, perhaps allowing that Aristotelian and computationalist arguments mutually reinforce one another, but concluding that they should simply all be thrown out together. Indeed, Searle explicitly maintains that not only computation, but also function and teleology more generally, are all observer-relative. Of biological phenomena, he writes:

> Darwin's account shows that the apparent teleology of biological processes is an illusion. It is a simple extension of this insight to point out that notions such as "purpose" are never intrinsic to biological organisms. . . . And even notions like "biological function" are always made relative to an observer who assigns a normative value to the causal processes. . . . In short, the Darwinian mechanisms and even biological functions themselves

---

59 Endicott, "Searle, Syntax, and Observer Relativity," 105.

60 Quoted in Luciano Floridi, ed., *Philosophy of Computing and Information: 5 Questions* (Copenhagen: Automatic Press, 2008), 16.

61 John Wilkins, "Information is the new Aristotelianism (and Dawkins is a hylomorphist)," *Scientia Salon*, May 1, 2014, at http://scientiasalon.wordpress.com/2014/05/01/information-is-the-new-aristotelianism-and-dawkins-is-a-hylomorphist.

> are entirely devoid of purpose or teleology. All of the teleological features are entirely in the mind of the observer.[62]

Searle would also deny that there is any level of physical reality that can be described accurately only from the functional point of view represented by Dennett's "design stance," as opposed to the purely causal level represented by the "physical stance." Searle writes:

> Where functional explanations are concerned, the metaphor of levels is somewhat misleading, because it suggests that there is a separate functional level different from the causal levels. That is not true. The so called "functional level" is not a separate level at all, but simply one of the causal levels *described in terms of our interests*. . . . When we speak of . . . functions, we are talking about those . . . causal relations to which we attach some *normative* importance. . . . [But] the normative component . . . [is in] the eye of the beholder of the mechanism.[63]

Now, there are several problems with this. For one thing, there are different respects in which biological phenomena might seem to exhibit teleology. The adaptation of an organism to its environment is one apparent instance of biological teleology. Developmental processes, and in particular the fact that some growth patterns are normal and others aberrant, are another. As several writers have pointed out, while Darwinism might explain away the first sort of example, it does not follow (contra Searle) that it explains away the second.[64] For another thing, the Scholastic would argue that it is a confusion to suppose that one can entirely replace teleological explanations with causal ones because even the simplest causal regularity will itself *presuppose* teleology. Again, if *A* regularly generates *B* rather than *C* or *D* or no effect

---

[62] Searle, *The Rediscovery of the Mind*, 51–52.

[63] Ibid., 237–238.

[64] Andre Ariew, "Platonic and Aristotelian Roots of Teleological Arguments," in *Functions: New Essays in the Philosophy of Psychology and Biology*, ed. Andre Ariew, Robert Cummins, and Mark Perlman (Oxford: Oxford University Press, 2002); Ariew, "Teleology," in Hull and Ruse, *The Cambridge Companion to the Philosophy of Biology*, 160–181; Marjorie Grene, "Biology and Teleology," in *The Understanding of Nature: Essays in the Philosophy of Biology* (Dordrecht: D. Reidel, 1974); and J. Scott Turner, *The Tinkerer's Accomplice: How Design Emerges from Life Itself* (Cambridge, MA: Harvard University Press, 2007).

at all, that can, according to the Scholastic, be only because generating *B* is the outcome toward which *A* is inherently directed as toward a final cause. If we do not recognize such rudimentary teleology, we will be stuck with Humean skepticism about causality.

These are large issues that cannot be settled here, and I have, in any case, argued for the reality of Aristotelian immanent teleology at length elsewhere.[65] It suffices for present purposes to note another problem with Searle's position. If we say that there is no teleology inherent in mind-independent reality, we are pretty clearly left with two options. We could, on the one hand, say that there is no teleology *at all*, *anywhere*, not even in the mind. That would be an eliminativist position, and it would be difficult at best to make such a position coherent. For, if there is no teleology or "directedness" of any sort, then there would be no "directedness" of the kind associated with the intentionality of thought. And it is notoriously difficult to deny the existence of intentionality coherently, since the very denial is itself a manifestation of intentionality.

Certainly Searle is no eliminativist about intentionality,[66] nor, it seems, about teleology or "directedness" in general. His view would seem to be the second alternative, according to which there *is* teleology *in the mind* even if there is none in mind-independent reality. But this would seem to entail either Cartesian dualism or property dualism, with all their associated problems—the interaction problem, epiphenomenalism, and so forth. To be sure, Searle claims not to be a dualist,[67] but like other critics, I find it hard to see how his view differs from dualism except verbally. Consider what Searle says about the sort of "directedness" associated with the intentionality of the mental:

> Intentional notions are inherently normative. They set standards of truth, rationality, consistency, etc., and there is no way that these standards can be intrinsic to a system consisting entirely of brute, blind, nonintentional causal relations. There is no normative component to billiard ball causation.[68]

---

65 Edward Feser, "Between Aristotle and William Paley: Aquinas's Fifth Way," *Nova et Vetera* (English) 11 (2013): 707–749; see also Feser, *Scholastic Metaphysics*, chapter 2.

66 Searle, *The Rediscovery of the Mind*, 6.

67 John R. Searle, "Why I Am Not a Property Dualist," in *Philosophy in a New Century*, 152–160.

68 Searle, *The Rediscovery of the Mind*, 51.

This is said in the context of his remarks about the absence of teleology from biological phenomena. The clear implication is that the human body and brain consist "entirely of brute, blind, nonintentional causal relations" of the "billiard ball" type, whereas the mind is the seat of the intentionality, rationality, normativity, and so on that cannot be "intrinsic" to the body and brain thus understood. That sounds pretty close to the Cartesian dichotomy between matter conceived of as pure mechanism devoid of thought and mind conceived of as pure thought irreducible to mechanism.

It is true that Searle regards the mental as *caused by* the physical, but then Cartesian substance dualists and property dualists also often affirm a causal relation between the physical and the mental. And while these dualists have famously had difficulty in explaining exactly *how* this causal relation works, Searle too admits that:

> We don't have anything like a clear idea of how brain processes, which are publicly observable, objective phenomena, could cause anything as peculiar as inner, qualitative states of awareness or sentience, states which are in some sense "private" to the possessor of the state.[69]

I discuss Searle's relationship to property dualism at greater length elsewhere,[70] and in any case, these are issues that obviously cannot be settled here. Suffice it to say that, from the Aristotelian-Scholastic point of view, the intractability of the debate between materialism and Cartesian forms of dualism is a consequence of what they have in common, namely, the "mechanistic" conception of nature that supplanted the Aristotelian-Scholastic conception—a conception that leaves no place for the teleological, and thus no place for the intentional. Like the Aristotelian, Searle has been critical of both materialism and Cartesianism, but from the Aristotelian point of view, Searle's own position is unstable, threatening to collapse back into one or the other of these alternatives precisely because he is also committed to the same "mechanistic" picture that they are. The key is to reject that picture and return to the one it supplanted. Contemporary computationalism, for all the flaws in it rightly identified by Searle, has the

---

69 John R. Searle, *The Mystery of Consciousness* (New York: The New York Review of Books, 1997), 8.

70 Edward Feser, "Why Searle *Is* a Property Dualist," in *Neo-Scholastic Essays* (South Bend, IN: St. Augustine's Press, forthcoming).

merit of gesturing precisely in the direction of such a return, however inadvertently.

Theological Implications

Let me close with some brief remarks on the theological implications of the position I have been defending. The first thing to note is a theological lesson I think should *not* be drawn from what I have been saying. Someone attracted to "Intelligent Design" theory might suppose that a good way to split the difference between Searle and the computationalist would be to agree with the computationalist that there really is computation in nature, but to agree also with Searle that all computation is observer-relative or conventional, rather than intrinsic, and then suggest that God is the observer who assigns a computational interpretation to natural objects so as to make of them computers. This would essentially be an updating of the conception of God's relationship to the world introduced by Descartes, Newton, and Paley, according to which the world is a kind of machine and God a kind of machinist. Or to go back to my Tarzan analogy, God, on this view, is like Tarzan and the world like the hammock he makes out of liana vines. On this view, the distinctions between substantial forms and accidental forms and between intrinsic and extrinsic teleology are dissolved. Even natural objects like liana vines would have merely accidental forms and extrinsic teleology imposed on them by God, and algorithms and software in nature would also be conceived of as accidental forms and instances of extrinsic teleology.

Now, from an Aristotelian-Scholastic point of view, though this sort of approach has been very common since Paley and others developed the modern "design argument," it is, metaphysically and theologically speaking, just a complete muddle. Natural objects are not a kind of artifact, and hence God's relationship to them is not that of an artificer. For it makes no sense to suggest that natural objects have only accidental forms and extrinsic teleology, since having an accidental form presupposes materials already having substantial forms and extrinsic teleology presupposes intrinsic teleology. Again, the hammock in our example has the *externally imposed* form and function it does only because the liana vines out of which it is made have the *intrinsic* tendencies they do just by virtue of being liana vines. So, hammock-making, watchmaking, and other kinds of artifice are simply not good models for understanding God's creation of the natural world, and thus thinking of God as a kind of electrical engineer or computer programmer is not a good model either.

Then there is the fact that making computation extrinsic to natural phenomena would entail a kind of occasionalism. Just as it is not a pocket calculator that really does arithmetic, but rather the user who does arithmetic by means of the calculator, so too it would not be the brain that processes information, but God who processes it using the brain the way you use your laptop. It would not be the genome that causes the embryo to develop the way it does, but rather God who does it using the genome the way an engineer might run a computer simulation, and so forth. What the Scholastic calls "secondary causality" would disappear from nature, and only God alone would be doing anything, as in the occasionalist theology of Malebranche.

That is not to say, however, that *intrinsic* teleology of the Aristotelian sort has no theological relevance. Far from it. Aquinas's "Fifth Way" of proving the existence of God begins with an argument for intrinsic teleology. Where Paley's "design argument," "Intelligent Design" theory, and the like go wrong is not by appealing to teleology, but rather in supposing that natural objects have teleology *in the way* that artifacts do, in supposing that reasoning from natural teleology to God has to do with weighing the probability that a complex natural structure could have come about via impersonal processes, in looking for gaps in current naturalistic explanations, and so forth. Aquinas's Fifth Way has nothing to do with any of that.

This is not the place to get into that subject, and I have defended the Fifth Way and explained how it differs from the "design argument" at length elsewhere.[71] It suffices for present purposes to note that, given its implicit Aristotelianism, the computationalist approach provides Thomists and other Aristotelians and Scholastics with conceptual and terminological resources by which contemporary naturalists might be made to understand and see the power of Thomistic, Scholastic, and Aristotelian arguments in natural theology. It might help them to explain both how the conception of nature on which traditional Scholastic natural theology was built is no pre-modern relic but is still defensible today, and how radically it differs from the conception of Paley and "Intelligent Design" theorists, whose arguments naturalists understandably regard as weak.

The computationalist approach may also help to make traditional natural law theory intelligible to contemporary naturalists. Traditional natural law arguments presuppose an Aristotelian metaphysics

---

71 Edward Feser, *Aquinas: A Beginner's Guide* (Oxford: Oneworld Publications, 2009), 110–120; see also Feser, "Between Aristotle and William Paley: Aquinas's Fifth Way."

of formal and final causality, as they presuppose that our faculties are of their nature directed toward certain ends. Goodness as an objective feature of the world is defined in terms of the realization of the ends toward which a thing is by nature directed, and badness in terms of the frustration of those ends. *Moral* goodness or badness enters the picture when a rational animal freely chooses to act in a way that either facilitates or frustrates the realization of the ends inherent in its nature.[72]

Now, these sorts of arguments become unintelligible on a mechanistic conception of nature because, on such a conception, the natural world is, as Searle puts it, "a system consisting entirely of brute, blind, nonintentional causal relations," and as he rightly notes, teleology and normativity cannot be intrinsic to such a system. But computational notions *are* normative. If a system is running a certain program, then there is a distinction to be made between "bugs" and "features" of the program, of the system functioning properly or malfunctioning, and so forth. Hence, if a naturalist regards some natural system as inherently computational, then he has at least implicitly affirmed that there is teleology and normativity in nature, after all, and thus he has conceded the basic metaphysical presupposition of traditional natural law theory.

All of this cries out for further development, of course. However that goes, what Searle and the Aristotelian can agree on is that the computationalist conception of nature is far more metaphysically loaded than most of its defenders realize. N&V

[72] See Feser, *Aquinas*, chapter 5; see also Feser "Being, the Good, and the Guise of the Good," in *Neo-Aristotelian Perspectives in Metaphysics*, ed. Daniel D. Novotny and Lukas Novak (London: Routledge, 2014), 84–103.