Real Algorithms: A Defense of Cognitivism

John Bolender
Columbia University (N.Y.) - Bilkent University (Ankara)

Cognitivism is the view that cognition, emotion, and even perception consist of algorithmic transformations of symbols. Since virtually all cognitivists are materialists, few of them believe that these algorithmic transformations of symbols take place in any spiritual or non-physical medium. So, for all practical purposes, cognitivism implies that the brain is an algorithmic symbol-transformer, i.e. a computer. On the cognitivist view, human thought supervenes upon the computational features of unconscious brain processes.

John Searle attacks cognitivism at its very root.[1] According to Searle, computation is not an objective feature of any consciously inaccessible brain process. Searle admits that any brain process can be described *as though* it were performing a computation but insists that such a description does not correspond to any objective, intrinsic feature of the process. According to Searle, unconscious brain processes are not intrinsically performing computations.

Can There Be Intrinsic Computation in Unconscious Nature?

In order to understand this claim, one must appreciate Searle's distinction between observer-relative and intrinsic features of reality. An observer-relative feature depends upon the attitude of observers and so could not exist if there were no sentient beings. By contrast, an intrinsic feature does not depend upon the attitude of observers and so could exist even in the absence of sentient beings. Consider a beautiful painting. The painting has some intrinsic features, such as its being composed of atoms and molecules, features in no way requiring the existence of conscious observers. But the painting's beauty is observer-relative since it is only beautiful in virtue of people judging

[1]l. John Searle, The <u>Rediscovery of the Mind</u> (Cambridge, Massachusetts: The MIT Press, 1992), Chapters 9 and 10.

it to be so. Accordingly, the painting would not have this feature if there were no sentient beings to observe it.

Scientific explanations exclusively appeal to intrinsic features of things such as mass, velocity, charge, and shape. This is due to the fact that scientists seek causal explanations of phenomena, and the observer-relative features of a thing do not account for any of its causal powers. The Mona Lisa has its causal powers regardless of whether or not there are any observers around to find it beautiful.

According to Searle, our brains are not computers because computation is not intrinsic to physics. On Searle's view, no unconscious process is ever intrinsically computational but only computational, if at all, relative to an observer. It follows that computational features, just like being beautiful, have no scientific explanatory efficacy and hence that cognitivism is bankrupt as a scientific research program.

Before evaluating this claim, one must be clear about what it is for a process to be computational. A process is computational just in case it follows an algorithm for transforming symbols. The classic example, of course, is a Turing machine whose semantic features, symbols inscribed on tape, change according to an algorithm as represented by the machine table. So, a necessary and sufficient condition for a process being computational is that it follow an algorithm for transforming its semantic features. Hence, the following two conditions are both necessary in order that a process be computational:

(i) The process has features with semantic content.
(ii) The process follows an algorithm.

Searle claims that no deep unconscious process, that is no process which is necessarily unconscious, can have intrinsic semantic properties or be intrinsically algorithmic. That is, no deep unconscious process ever satisfies (i) or (ii) in virtue of its intrinsic features alone. For example, no brain process can be intrinsically intentional unless it is also capable of being conscious. Furthermore, no unconscious brain process can intrinsically go through the steps of some algorithmic operation unless it is at least potentially-conscious. Since both of these claims must be refuted if one is to defend cognitivism, the challenge for the cognitivist is to show that (i) and (ii) are satisfied by some deep unconscious processes wholly in virtue of their intrinsic properties. Only then can the processes be said to be intrinsically computational.

In this paper, I will focus almost exclusively on the question of whether (ii) can be satisfied by intrinsic properties. The challenge of showing that (i) can be satisfied wholly by intrinsic properties is, of course, the challenge of naturalizing content. Since so many philosophers are writing on that issue, I do not feel that it is necessary to address (i) in this paper, but there can be no complete response to Searle until it is shown that content can be naturalized.

Algorithmicity, a Necessary Condition for Computation

What does it mean for a process to be algorithmic? Dictionary definitions imply that an algorithm is a procedure, decomposable into steps, which guarantees or probabilifies a certain result or end-state. So in order for a process to be algorithmic, its behavior must at the very least be in agreement with such a procedure, i.e. it must be describable as though it were following the algorithm. In fact, such mere agreement has often been taken as a sufficient condition for a process being algorithmic, but this constitutes an extremely permissive criterion for determining whether a process is algorithmic, for any physical process whatsoever is in agreement with some algorithm or other. In order for a process to so agree, it is enough that it be describable as following a series of steps eventuating in some guaranteed (or at least probabilified) end-state. Furthermore, all it takes for something to be describable as following a series of steps is for it to be decomposable into simpler sequential processes, and all it takes to guarantee or probability a certain end-state is for there to be laws of nature. So, for example, a stone accidentally rolling down the side of a mountain behaves in agreement with an algorithm since any stage in that process can be described as though it were a step in some procedure terminating in a guaranteed or probabilified end-state, say, arriving at the final resting place of the stone at the mountain's foot. If behaving in agreement with an algorithm is sufficient for following an algorithm intrinsically, then surely all processes are intrinsically algorithmic.

Daniel Dennett agrees that any process whatsoever behaves in agreement with some algorithm or other and concludes that all processes are algorithmic. He remarks, "So what?"[2] The problem with this lenient view is that any process with changing semantic features would automatically qualify as following a symbol-manipulating algorithm, i.e. as being computational. In fact, if computation is sufficient for intelligence, as the cognitivist claims it to be, any process with changing semantic features would automatically count as intelligent, and this is too generous.

Imagine that philosophers, cognitive psychologists, and neuroscientists all come to agree on some plausible theory of naturalized semantic content such that one can discover intrinsic semantic features in nature. If a cognitivist were to adopt Dennett's permissive view of algorithms, they would have to admit that it would be inconceivable for, say, neuroscientists to discover changing semantic features in a physical process, such as a brain, without also

2. Daniel C. Dennett, Darwin's Dangerous Idea (New York: Simon & Schuster, 1995), p. 59.

recognizing that very process to be computational and hence intelligent. But surely this is quite conceivable. One could imagine a process with semantic features so chaotic and random in their transformations as to be obviously non-computational and non-intelligent - a bit like the accidental semantic transformations one might find in a bowl of alphabet soup. We need some way of separating the wheat from the chaff, the genuinely algorithmic from the merely causal.

In contrast to Dennett, Searle believes that one can distinguish processes in nature which are algorithmic from those which are not, but the difference, or so he claims, is purely a matter of interpretation.[3] According to Searle, one sees certain processes as following algorithms (e.g., the heart follows an algorithm in causing blood circulation), and one sees other processes as nothing more than instances of brute causation (e.g., the lightning causes the thunder but without following any algorithm in doing so), but there is no intrinsic feature of algorithmicity possesssed by the former and not the latter. We simply interpret the former as rule-governed or goal-directed and the latter as nothing but an instance of brute causation. This view of algorithms in unconscious nature, however, is as unacceptable to the cognitivist as Dennett's. Since following an algorithm is a necessary condition for being computational, Searle's view would imply that computation in unconscious processes is always observer-relative, never intrinsic, and hence scientifically worthless.

Given that any process whatsoever behaves in agreement with some algorithm or other, the cognitivist would appear to be between a rock and a hard place. If, as Dennett suggests, following an algorithm is nothing but behaving in agreement with that algorithm, then all processes are algorithmic. But this would imply that any process with semantic features is automatically computational and hence (if cognitivism is correct) automatically intelligent, an absurd claim. That is the rock. The hard place is Searle's view that algorithmicity, and hence computation as well, is observer-relative, a mere matter of interpretation. Can the cognitivist escape this seeming dilemma, this apparent choice between pan-algorithmicity and observer-relative algorithmicity?

Functions as a Ground for Algorithmicity

One can begin to escape the apparent dilemma by considering what criteria must be met by an artifact in order for it to follow an algorithm. In virtue of what, for example, does a carburetor's behavior follow an algorithm for producing an inflammable vapor? Part of an answer consists in pointing out that the carburetor's behavior agrees with an algorithm for mixing fuel

3. Searle, op. cit.

with oxygen resulting in the production of such a vapor, but we have already seen that this is not a sufficient condition. There must be some further requirement to be met for it to follow this algorithm, and what could this requirement be other than its mixing fuel and oxygen *in order to* produce an inflammable vapor? If it were simply an accident that this object mixes fuel and oxygen then, even though its behavior would be in agreement with such an algorithm, it would not genuinely be following it. I submit that, in addition to being in agreement with a given algorithm (A), a process genuinely follows A only when it behaves in agreement with A *in order to* realize A's end-state, that is, when it is teleologically directed toward that state. (This agrees with some dictionary definitions of "algorithm," according to which an algorithm is a step-by-step procedure "for accomplishing some end.")

For example, on this view, the heart intrinsically follows a blood-pumping algorithm partly because its behavior agrees with such an algorithm but also because it is a function of the heart to produce the end-state of the algorithm, namely the circulation of blood. By contrast, the heart does not intrinsically follow a make-thumping-sounds algorithm because, even though its behavior agrees with such an algorithm, it is no function of the heart to produce such sounds.

The cognitivist can say that any deep unconscious brain process intrinsically follows an algorithm if its behavior is in agreement with the algorithm, and one of its functions is to realize the end-state of that algorithm. Furthermore, the process is following a symbol-manipulation algorithm if the process has semantic features and it is one function of the process to transform those features. To follow such an algorithm intrinsically is to be intrinsically computational. So a deep unconscious brain process is intrinsically computational if it has intrinsic semantic features which change according to some algorithm and it is a function of the process to bring about the end-state of that algorithm.

So an appealing response to Searle's scepticism about algorithmicity is to say-that a deep unconscious process intrinsically follows a certain algorithm if

(1) the process behaves in agreement with the algorithm,

and

(2) one function of the process is to produce the end-state of the algorithm.

Prima facie, this allows for intrinsic algorithmicity without succumbing to Dennett's view that all processes are algorithmic, since not just any cause has the function of producing its effect. But Searle anticipates this move and thoroughly beats up on it by claiming that function ascriptions are themselves observer-relative.

Searle writes that

> ... when we say that the heart functions to pump blood, the only facts in question are that the heart does, in fact, pump blood; that fact is important to us, and is causally related to a whole lot of other facts that are also important to us, such as the fact that the pumping of blood is necessary to staying alive. If the only thing that interested us about the heart was that it made a thumping noise or that it exerted gravitational attraction on the moon, we would have a completely different conception of its "functioning" and, correspondingly, of heart disease. To put the point bluntly, in addition to its various causal relations, the heart does not have any functions. When we speak of its functions, we are talking about those of its causal relations to which we attach some *normative* importance.[4]

According to Searle, functions are observer-relative because function ascriptions presuppose values which are themselves in the eye of the beholder.

> We do indeed "discover" functions in nature. But the *discovery* of a natural function can only take place within a set of prior *assignments* of value ... It is because we take it for granted in biology that life and survival are values that we can discover that the function of the heart is to pump blood. If we thought the most important value in the world was to glorify God by making thumping noises, then the function of the heart would be to make thumping noise, and the noisier heart would be the better heart. If we valued death and extinction above all, then we would say that a function of cancer is to speed death... As far as nature is concerned intrinsically, there are no functional facts beyond causal facts. The further assignment of function is observer relative.[5]

I take Searle to be saying that a statement of the form "The function of F is to G" is equivalent to "F is causally responsible for G, and G is good." So, for example, to say that the heart functions to pump blood is tantamount to saying that the heart is causally responsible for the pumping of blood, and the pumping of blood is good. This implies that functions are observer-relative if values are observer-relative.

Clearly, in insisting that values are not intrinsic, Searle is assuming a sturdy fact/value distinction in the spirit of Hume. On the Humean view, rather than actually discovering value in the world, the mind superimposes value upon it in accordance with its own desires. Values are observer-relative on this view since they depend upon thinking about the world in terms of

4. Ibid., p. 238.
5. John Searle, The Construction of Social Reality (New York: The Free Press, 1995), pp. 15-6.

one's desires, and so if functions presuppose norms then they too are observer-relative rather than intrinsic.

If this is the correct analysis of function ascriptions, then, contrary to the hopes of the cognitivist, condition (2) does not place any constraint whatsoever on condition (1). Assuming that a process satisfies (1) in virtue of behaving in agreement with some algorithm, then the production of the end-state of that algorithm can indeed be construed as the function of the process relative to some set of values or other. Obviously, on Searle's view, grounding algorithmicity in functions renders algorithmicity observer-relative, for functions themselves are observer-relative. Even if semantic features of unconscious processes are intrinsic, the cognitivist would still be forced to admit that computation in unconscious processes is never intrinsic as long as their algorithmicity is observer-relative.

In fact, we can now better understand why Searle considers algorithmicity to be observer-relative. He agrees that a process is following an algorithm if it behaves in agreement with the algorithm and has the function of bringing about the algorithm's end-state, but, if functions are observer-relative, then it follows that algorithmicity is observer-relative as well, a matter of value judgment.

Purely Descriptive vs. Normative Theories of Functions

But is Searle's normative analysis of function ascriptions the right analysis? Larry Wright has proposed an analysis which is purely descriptive, free of any normativity whatsoever.[6] According to Wright, the heart functions to pump blood because, both, the heart tends to pump blood, and the heart has come to be because of its tendency to pump blood. In general terms, the function of any x is F just in case (A) x tends to bring about F, and (B) x has come to be by virtue of its tendency to bring about F. Since Wright attempts to analyze functionality wholly in terms of causation, his is known as an *etiological* analysis of functions.

However, stated in this form, this etiological account is subject to counterexamples. Consider, for example, a stick pinned to a rock in a stream.[7] The stick is pinned to the rock in virtue of a backwash which is itself formed in virtue of the stick's being pinned to the rock. The stick originally came to be pinned to the rock purely by chance, but once contact was made it continued to be pinned to the rock due to the resultant backwash. In this case, we have a condition, the stick's being pinned to the rock, which produces an effect, the backwash, which itself causally explains why the stick remains

6. Larry Wright, "Functions" The Philosophical Review (1973), pp. 137-68.
7. See Mark Bedau, "Can Biological Teleology Be Naturalized?" The Journal of Philosophy (1991), pp. 647-55, who attributes the counterexample to Robert Van Gulick.

pinned to the rock. Since both conditions (A) and (B) are satisfied, according to Wright's analysis, the function of the stick's being pinned to the rock must be to produce a backwash. However, as should go without saying, it is very counterintuitive to suppose that a stick's being pinned to a rock in a stream would have any function at all unless it were deliberately placed there by an intelligent being.

This particular counterexample can be avoided by strengthening the etiological account of functions by an appeal to natural selection or *differential reproduction* (to use a less anthropomorphic term). On this fortified version, the etiology referred to in B must be spelled out in terms of differential reproduction. For example, the heart's function is to pump blood because (A) it tends to pump blood and (B) creatures with hearts have flourished because the circulation of blood increases the chances of surviving to reproductive maturity.

However, Bedau has convincingly argued that even if the conjunction of A and B provides a necessary condition of functionality, it does not provide a sufficient condition. What must be added is a claim that the relevant effects of, say, the heart are valuable. So, if Bedau is correct, ascriptions of function once more turn out to be normative, and Searle would appear to be vindicated.

Bedau illustrates his point by discussing differential reproduction among non-living things, specifically clay crystals.[8] Lifeless populations of clay crystals exhibit all four individually necessary and jointly sufficient conditions for differential reproduction; namely reproduction, variation, heredity, and adaptivity. A given population of clay crystals can indeed reproduce more successfully than other populations because of some heritable feature of its crystalline structure, call it 'D.' Such a feature, although initially a mutation, can proliferate because it aids in reproductive success. D may aid in such success, for example, by tending to dam up streams thereby turning them into shallow pools. During the dry season, these pools evaporate leaving clay dust which is carried by the wind to other streams thus beginning the cycle anew. The result is a proliferation of crystal populations with feature D. Hence, if the heart functions to pump blood by virtue of this effect enabling the creature to survive to reproductive maturity, then crystal feature D must function to dam up streams since its having this effect enables a certain population of clay crystals to reproduce more efficiently and thereby perpetuate D.

Bedau appeals to our intuitions. According to him, it seems strongly counterintuitive to say that crystal feature D has any function whatsoever,

8. Bedau cites Richard Dawkins for this example. See Dawkins, The Blind Watchmaker (New York: Norton, 1986), p. 153.

including the function of damming up streams, even though the causal ancestry of D satisfies the requirements posited by the etiological theory. This is in sharp contrast to our intuitions about other products of differential reproduction such as hearts and lungs. According to Bedau, the relevant difference between, say, D and the heart is that the effects of D are not good for anything while the heart's pumping blood is good for the organism in question. The world of biology, or so Bedau claims, is a world containing values. By contrast, the lifeless world of clay crystals harbors no value and, accordingly, harbors no functions. It is an essential condition for something being functional, according to Bedau, that it brings about that which is good.

So Bedau concludes, quite correctly I believe, that any function ascription implies a value judgment. Bedau further reasons that since there are intrinsic functions in nature and since functions require values, there must be intrinsic values in nature. But this modus ponens inference could just as easily be replaced by a modus tollens inference to the effect that since there are no intrinsic values in nature then there can be no intrinsic functions either. I consider the modus tollens inference to be preferable since it is less radical, I believe, to acknowledge a world without intrinsic functions than to acknowledge a world with intrinsic values. Physicists have lived in the former world for centuries, whereas acknowledging the latter would violate a fundamental assumption of all the modern sciences, namely the fact/value distinction. So Bedau's thought experiment, despite his own intentions, supports Searle's claim that functions are observer-relative.

Revising or Replacing the Folk Concept of Function

In reading Searle and Bedau on functions, one might be left with the impression that the cognitivist faces the following dilemma: Either there are no intrinsic functions in unconscious nature, in which case there is no intrinsic computation in unconscious nature (unless one espouses Dennett's overly permissive view); or there are intrinsic functions in unconscious nature, in which case there are also intrinsic values in unconscious nature. The former horn of the supposed dilemma would clearly refute a scientific cognitivism, for cognitive scientists require that the features they study be intrinsic, not observer-relative. The second horn of the supposed dilemma, the claim that there are intrinsic values in unconscious nature, although perhaps acceptable to some cognitivists, would be a step too radical for many scientists and philosophers to take, since the fact/value distinction is one of the most deeply entrenched assumptions of modern science.

But this supposed dilemma, I submit, is not genuine. The problem with both Searle and Bedau's discussions is that they hold up our intuitive, pre-scientific folk concept of function as the analysandum for any theory of functions.[9] So when Searle and Bedau discover that functions require values, they are discovering something about our old, pre-scientific conception of functions. Assuming a sturdy fact/value distinction, Searle and (inadvertently) Bedau have shown that functions do not exist intrinsically in unconscious nature, but this is only to show that an old, pre-scientific concept of ours lacks any intrinsic application to unconscious nature. This is surely no great surprise. Many of our old, pre-scientific folk concepts lack application to the world as conceived by the contemporary sciences. For example, our pre-scientific, folk concept of up-down has no intrinsic application to the universe as described by contemporary astrophysics and cosmology, a universe in which all directions are equal.

Scientific progress sometimes requires that an old concept be discarded in favor of a new one. I submit that we should acknowledge the implicit normativity in function ascriptions but replace the folk notion of function with a new, purely descriptive concept. This new concept would enable us to recognize intrinsic algorithmicity in the unconscious natural world and, with the help of a theory of naturalized content, would enable us to identify intrinsic computations in unconscious processes.

Here is an example of such a replacement: the word "mass" in the context of Newtonian mechanics does not refer to the same concept as does the word "mass" in the context of Einsteinian mechanics.[10] Newtonian mass is a property whose value for an object remains constant relative to all frames of reference. By contrast, Einsteinian mass is a property of an object which varies in value depending upon reference frame, taking a different value subject to the velocity of the frame of reference relative to the object whose mass is being measured.[11]

It is natural to say that the concept of Einsteinian mass replaced that of Newtonian mass, but what does it mean to say that? It must mean that scientists use the concept of Einsteinian mass to do the same explanatory work that the concept of Newtonian mass was meant to do. For example, where scientists may have once attempted to explain the depth of a lunar crater by attributing a certain Newtonian mass to the meteoroid responsible

9. Mohan Matthen also faults Bedau on this point. See Matthen, "Naturalism and Teleology" The Journal of Philosophy. (November 1991), pp. 656-7.

10. In using this example, I am most emphatically NOT taking a stand on the debates over whether meaning is holistic or whether competing scientific theories are ever incommensurable. Nor do I believe that these debates have any direct bearing on the question of whether there are intrinsic algorithms in nature. I do, however, insist that conceptual replacement occurs in the sciences at least sometimes and that "mass" provides a good example of such.

11. See Patricia Churchland, Neurophilosophy (Cambridge, Massachusetts: The MIT Press, 1986), pp. 286-7.

for the crater, they would now instead attribute to the meteoroid an Einsteinian mass in explaining the crater's depth. Generally speaking, using a new concept to do all of the explanatory work that an older concept was meant to do is tantamount to replacing the old with the new. So I propose that if a purely descriptive concept can be shown to be capable of all the explanatory work of which our intuitive, normative concept of function is capable, then scientists should use the former descriptive concept in place of the latter normative one.

But this raises the question of how one judges that one concept is equal in explanatory power to another. To what criteria can one appeal in deciding whether some descriptive concept is capable of all the explanatory work of which the normative notion of function is capable? I wish to emphasize that one only needs to know conditions which are sufficient for the one concept to have explanatory power equal to the other. We are not searching for some general theory of when two concepts are equal in explanatory power but are only concerned with showing that some particular descriptive concept is capable of the same explanatory duty as the normative notion of function. Hence, only the requisite sufficient conditions are needed.

What might such sufficient conditions be? In answering this question, let us use some convenient terminology. Let "function-statement" refer to any statement ascribing at least one function to something. Examples of function-statements include "Tropism in plants functions to maximize their exposure to light," "The heart functions to pump blood," and "This neural process functions to transform symbols." Let "D" be any purely descriptive concept, and let a "D-statement" be any assertion attributing D-ness to something. For example, if D is the concept of being square in shape, then "That figure is square" would be a D-statement. This terminology permits us to re-state the question about the explanatory powers of concepts in terms of the explanatory powers of statements: The question of whether some purely descriptive concept can replace that of function is the question of whether there is a concept D such that the D-statements are capable of doing the same explanatory work as function-statements.

For some specific D-statement, how does one know that it has the same explanatory power as some function-statement? I submit that if the D-statement plays the same role in explanation that the function-statement plays, then it is equal in explanatory power to that function-statement. But what does it mean "to play the same role in explanation"? Part of what it means is that the D-statement can be used to explain the same explananda as does the function-statement. For example, if the function-statement "The heart functions to pump blood" can be offered as an explanation of why humans have hearts, then if offering the D-statement also suffices as an explanation of why humans have hearts, then the D-statement is capable of playing at least some of the same explanatory role as the function-statement.

Unfortunately, however, what it means to play the same explanatory role cannot be explicated wholly in terms of explaining the same explananda. Lawlike generalizations also play a role in functional explanations, and so in order to play the same explanatory role as some function-statement, the pertinent D-statement must not only be capable of explaining each explanandum which the function-statement explains but must do so via reference to the same lawlike generalization. If the D-statement explains all the explananda explainable by citing the function-statement and if it does so in each case by reference to the same generalization as would appear in the functional explanation, then one can rest assured that the D-statement has the same explanatory efficacy as does the function-statement.

Please bear in mind that this is only an attempt to state a <u>sufficient</u> condition for one statement or concept performing the same explanatory work as another. It is conceivable that there are cases in which one performs all the explanatory work of another even though they do not explain by appeal to the same generalizations. In fact, the earlier example of Newtonian versus Einsteinian mass would be a case in point, since an explanation in terms of Einsteinian mass would appeal to Einsteinian, not Newtonian, generalizations without being any less explanatory. But a sufficient condition is all I need in order to show that some purely descriptive concept is capable of doing all the explanatory work of the concept of function.

Now we can return to that very question, namely whether there is a descriptive concept D such that D-statements can do the same explanatory work as function-statements. The point of the above discussion of explanatory power is to show that one can answer this question by considering which generalizations are appealed to in functional explanation. *If there is a concept D such that the D-statements explain the same explananda as do function-statements by appealing, in each case, to the same generalization as does the relevant function-statement, then D will have the same explanatory power as does the concept of function.*

So what sort of generalization figures in functional explanation? According to G.A. Cohen, such a generalization is used in explaining why a thing has F by showing that F increases the likelihood of its having some other property. To a rough approximation, any such generalization is to the effect that if an object is of such a nature that its being F increases the likelihood of its being E, then it will be F. To be more precise, such generalizations have the form

(G) IF it is true of an object o that if it were F at t_1, then it would, as a result, be E at t_2, THEN o is F at t_3.[12]

(An important qualification to note is that t_2 never precedes t_1, and t_3 never precedes t_2. This is to ensure that the explanation is not "teleological" in the scientifically unacceptable sense involving backwards causation.)

For example, in explaining why human beings have hearts, one might offer the functionalist explanation that the heart functions to pump blood. By offering this statement as an explanation, one appeals to the generalization that humans are of such a nature that IF it is the case that if their having hearts will increase the chances of their blood circulating, THEN they will have hearts.

As an example of such a generalization, Cohen suggests the following: Let o be the species cow, F = long-tailed, and E be the power to swish flies away. One thus arrives at the generalization that if cows are such that their being long-tailed enables them to swish flies away, then they will develop long tails. Another example of such a claim is that if clay crystals are such that their possessing feature D will result in their forming dams, then they will develop D. (Since these generalizations are obviously not strictly deterministic, Cohen suggests that one may add some probabilistic component to G or construe it as a law-sketch rather than a full-blown law.)

Now let us consider a case in which some function-statement can be used to explain, say, the possession of hearts by human beings. In explaining this phenomenon, one states that

> (N) The function of the heart is to circulate blood,

and one appeals, at least implicitly, to the generalization

> (G') If humans in the past have been such that their possessing hearts resulted in the circulation of their blood, then there is a greater probability than otherwise of their currently possessing hearts.

The conjunction of N and G' explains the possession of hearts by human beings because N implies that the antecedent of G' is satisfied. Such is the nature of functional explanation.

But here is the rub: N is not the only statement having this implication. Obviously, instead of N, one could have stated that

> (D) Humans in the past were such that their possessing hearts resulted in the circulation of their blood.

D obviously implies that the antecedent of G' is satisfied since D simply is the antecedent of G'. So there are purely descriptive causal assertions, such as D, which can explain the same phenomena as function-statements by appealing to the same generalizations, such as G'. So function-statements are

12. G. A. Cohen, Karl Marx's Theory of History: A Defence (Princeton, New Jersey: Princeton University Press, 1978), Chapter IX.

scientifically expendable in favor of certain purely descriptive causal assertions.

But not just any causal assertion can do the explanatory work of a function-statement, since not just any causal assertion is explanatory in conjunction with a generalization of form (G). Consider the following causal assertion:

(A) Lightning causes thunder.

Assume, what is surely false, that A can explain the occurrence of lightning when proffered in conjunction with some generalization of form G. If this were the case, the generalization would presumably be as follows:

(G") The past propensity of lightning to produce thunder probabilifies the present occurrence of lightning.

But G" is patently false. So not just any causal assertion can be put to the same explanatory use as a function-statement. That is, not just any causal assertion explains in conjunction with a law of form G. (This point will prove to be extremely important in silencing Searle's claim that any unconscious process is as intrinsically algorithmic as any other.)

Those causal assertions which can be put to the same explanatory use as function-statements ascribe "functions" *according to Wright's theory of functions*. Recall that, on Wright's view, "the function of X is Z" means that X is present (or occurs) because it does Z, and Z is an effect of X. I have already pointed out that Bedau has shown this to be an inadequate theory of our folk concept of function. However, Wright's discussion is still of great value, since his concept of "function" is capable of doing the same explanatory work as our folk concept of function. *Therefore, I propose that Wright's theory no longer be construed as an analysis of the concept of function but, instead, as a proposal for the adoption of a new concept to replace that of function.*

Let "function(d)" refer to any function as conceived by Wright, and let "function(n)" refer to any function as conceived by our folk concept of function. (The sense behind the nomenclature is that function(d)-statements are purely descriptive, while function(n)-statements are partly normative.) Examples involving the stick pinned to the rock in the stream and the evolution of adaptive features in clay crystals show that there are functions(d) which are not functions(n). However, statements ascribing functions(d) can be used to do all the explanatory work that statements ascribing functions(n) are used to do. This is because, when a function(n)-statement is used to explain, there is implicit appeal to a type-(G) generalization; but, in any such case, there is also some function(d)-statement which can explain the same explanandum by dint of the same generalization. For example, one can explain the presence of hearts in humans either by pointing out that they function(n) to pump blood or that they function(d) to pump blood.

To summarize a few of the points recently made and draw further conclusions: A function(d) of X is Z just in case X is present (or occurs) because it does Z, and Z is an effect of X. A function(n) of X is Z just in case Z is a function(d) of X, and Z is valuable. So, even though they do the same explanatory work, *function(d)* is a purely descriptive notion, while *function(n)* is normative. Accordingly, functions(d) are intrinsic features of phenomena, while functions(n) are observer-relative. Bedau's examples show that our pre-scientific folk concept of function is a concept of function(n), not function(d). However, since functions(d) do all the explanatory work performed by functions(n) in addition to being intrinsic, the concept of *function(d)* should replace that of *function(n)* in scientific discourse.

Furthermore, not just any effect of a phenomenon is a function(d) of that phenomenon. Lightning, for example, does not function(d) to produce thunder. To say, for another example, that the heart functions(d) to pump blood is not just to say that the heart brings about the pumping of blood. It is also to say that appealing to the heart's propensity to pump blood explains the presence of the heart in terms of a type-(G) generalization.

One corollary of this theory of functions(d) is that not only will many unconscious neural processes have functions(d), but many inanimate phenomena will as well. The stick pinned to the rock does function(d) to produce a backwash, and crystal feature D does function(d) to produce dams. But the charge of being counterintuitive is here avoided since the concept of *function(d)* is not meant to be the same as our folk concept of "function," which is, of course, the concept of *function(n)*. It is no charge against function(d)-statements that they conflict with our intuitions about functions(n). Just as "function(d)" is a term of art, the concept of function(d) is a concept of art. It is meant to be as explanatory as the older concept of function(n), but it is not meant to agree with our pre-theoretical intuitions about functions(n). Hence, one's intuitions about sticks and crystals in unconscious nature not having functions(n) does not extend to functions(d). So, using *function(d)* as a criterion of algorithmicity does not lead to Dennett's highly permissive pan-algoritnmicity.

It is noteworthy that the concept of *function(d)* avoids the Scylla/Charybdis of pan-teleologism vs. observer-relative teleology. Not just any causal assertion is a function(d)-statement. It is not a function(d)-statement to say that the heart exerts a gravitational force on the moon. Nor is it a function(d)-statement to say that the heart makes a thumping sound. Furthermore, whether a cause functions(d) to produce some effect is a purely intrinsic feature of the relevant phenomena.

Now we have a notion of an unconscious process in nature occurring *in order to* realize the end-state of an algorithm without utilizing any but purely descriptive concepts and without being forced to the reductio of having to say that all processes are intrinsically algorithmic. Any process, be it unconscious, inanimate or whatever, intrinsically follows an algorithm (A) if, and only if, it behaves in agreement with A, and it has the function(d) of producing the end-state of A. So, for example, the heart does not intrinsically follow an exert-gravitational-force-on-lhe-moon algorithm because, even though the heart's causal transactions are in accord with such an algorithm, it is not a function(d) of the heart to have this effect. However, crystal feature D does play a role in an intrinsic dam-making algorithm, since it is the function(d) of D to make dams. So, ascribing algorithmicity to unconscious natural processes is not a free-for-all in which all processes are equally intrinsically algorithmic. Therefore, in saying that certain unconscious brain processes are following certain algorithms, the cognitivist is not making a vacuous statement. We now have a value-free and selective criterion for determining which deep unconscious processes are algorithmic and which are not.

It is important to re-emphasize at this point that this is not meant as a complete account of intrinsic computation but only of intrinsic algorithmicity, a necessary but not sufficient condition of intrinsic computation. In helping to form dams, crystal feature D is intrinsically following an algorithm on my view, but it is not intrinsically computational since it has no intrinsic semantic features. In order to be computational, a process must follow a symbol-manipulating algorithm. So, for example, a neural process is intrinsically computational if it both has intrinsic semantic features and functions(d) to transform those semantic features. So this is only a partial response to Searle, since a complete response would not only require a naturalization of algorithmicity but a naturalization of semantic content as well.

Functions(d), of course, are in no way normative. The stick in the stream does no wrong if it suddenly breaks free of the rock to which it has been pinned. This shows that cognitivists speak carelessly in saying that intelligence consists in *"rule-governed"* manipulations of symbols. Since rules have no place in a scientific account of the world (given a rigid fact/value distinction), following an algorithm has for too long been equated with following a rule. I believe that one virtue of the present discussion is that it accounts for algorithmicity without appealing to any normative notions such as rule-following. Cognitive scientists will continue using the expressions "rules" and "rule-following." I am in no position to stop them. But I suggest that their use of such terms be interpreted, more charitably than literally, in a purely descriptive manner. For example, to agree with Noam Chomsky that language acquisition is to be understood in terms of a neural mechanism that follows certain "grammatical rules" is just to say that there are neural processes whose function(d) is to enable the child to learn only those languages sharing certain structural features. So one can grant that neural processes follow rules but only in the same purely descriptive sense in which the heart follows a blood-pumping rule. That is, what is called "rule-following" should charitably be construed as functioning(d) to produce a certain result.

I want to re-emphasize here that this is not to say that all effects are functions(d) of their causes. It is not to say that neural processes are "rule-governed" in the trivial sense in which the motions of the planets are "rule-governed." If cognitive scientists wish to continue using the misleading term "rule-governed," it should be construed as applying precisely to those processes which are intrinsically algorithmic, namely those processes which function(d) to produce their effects. The motions of the planets are not algorithmic in this sense, although many unconscious processes, from crystal dam-making to blood circulation, are. So when cognitivists say that semantic neural processes are "rule-governed," what they say, although perhaps misleading in wording, is not vacuous.

This purely descriptive characterization of algorithmicity also permits one to speak of brain "syntax" without appealing to norms. Syntax should be understood in terms of (intrinsic) symbol manipulating algorithms, not norms. If a neural process functions(d) to transform symbols in certain ways, then it would be a-syntactic for such a process to transform symbols in a way contrary to that function(d). However, an a-syntactic semantic brain process is not doing anything wrong. No norms are violated. It is simply behaving in a way contrary to its function(d). Therefore, given a purely descriptive account of what it is to be a symbol, a purely descriptive characterization of algorithmicity would give us a non-normative account of neural syntax.

Summary and Conclusion

The sciences are exclusively concerned with intrinsic features of reality such as mass and shape, not with observer-relative features such as goodness and beauty. Hence, in order for cognitivism to be a scientific hypothesis, it must be coherent to attribute intrinsic computational features to some unconscious processes. Hence cognitivism requires that some unconscious processes be intrinsically algorithmic, but on the face of it *all* processes are algorithmic in the trivial sense of behaving in agreement with some algorithm or other. This either threatens to make algorithmicity observer-relative or to make all processes algorithmic. Neither possibility is acceptable to the cognitivist.

In order to avoid pan-algorithmicity and observer-relative algorithmicity, one could appeal to teleology. The only truly algorithmic processes, on this proposal, are those which occur *in order to* achieve the relevant algorithm's end-state. However, even though this may avoid the Scylla of pan-algorithmicity, it falls to the Charybdis of observer-relative algorithmicity - at least if Searle is correct. For Searle claims that such teleological notions as function imply value judgments and, hence, do not correspond to intrinsic features of reality.

Bedau's thought experiments support Searle's claim, but one must bear in mind that Searle and Bedau are analyzing our pre-scientific folk concept of function. That normative concept could conceivably be replaced by a purely descriptive concept capable of all the explanatory work of the old. In fact, Wright's theory of functions, which was originally meant as an analysis of our vernacular concept, serves quite well as a scientifically acceptable *replacement* for that vernacular, normative concept. Wright's non-normative construal of functions, what I call "functions(d)," can be used as an objective criterion for distinguishing algorithmic from non-algorithmic processes.

But appeal to intrinsic algorithmicity is only a partial response to Searle. In order for a process to be intrinsically computational, it must not only be intrinsically algorithmic but intrinsically semantic as well. Therefore, a naturalization of semantic content is also essential for the cognitivist project. In this paper, I have only addresssed the issue of naturalizing algorithmicity, leaving the naturalization of content to others.

I conclude that appeal to both the notion of function(d) and a naturalization of content enables one to formulate purely descriptive and hence scientifically acceptable characterizations of computation, algorithmicity, and syntax. Provided that a naturalization of intentionality is forthcoming, it is coherent to explain cognition in terms of computation. Searle is wrong to suggest otherwise. However, whether this is the best explanation can only be decided by empirical testing.[13]

13. I would like to thank Sanford Goldberg and Shaughan Lavine for helpful comments on the paper.