CONSCIOUSNESS AND COGNITION 2, 320–333 (1993)

Comment on Searle: Philosophy and the Empirical Study of Consciousness

ANTHONY DARDIS

Philosophy/104 Heger Hall, 115 Hofstra University, Hempstead, New York 11550-1090

I make three points about Searle's philosophical work on consciousness and intentionality. First, I comment on Searle's presentation and paper "The Problem of Consciousness." I show that one of Searle's philosophical claims about the relation between consciousness and intentionality appears to conflict with a demand he makes on acceptable empirical theories of the brain. Second, I argue that closer attention to the difference between conceptual connections and empirical connections corrects and improves Searle's response to the so-called "Logical Connections" argument, the argument that claims that mental states cannot be causes, since they are conceptually connected with actions. Third, I give a formulation of his Chinese Room Argument that avoids some tempting responses. © 1993 Academic Press, Inc.

1. I want to make three points about Searle's philosophy of mind. First, I offer a comment and criticism of Searle's presentation at the Claremont Conference on Consciousness and Cognition (March 6, 1993), "The Problem of Consciousness," (Searle, 1993). I indicate a *prima facie* tension between two things Searle says about consciousness. The tension has to do with the boundary, vague as it is, between philosophy and empirical inquiry. If there really is an analytic/synthetic distinction (and I think there is), and there really are some things we can learn by conceptual analysis, then, equally, we cannot learn those things by doing empirical inquiry. The second point continues in this vein. Philosophers used to think the "logical connections" argument showed that reasons could not be causes. That argument is unsound, but there is a closely related, and sound, argument that shows (on plausible assumptions) that reason *properties* cannot be causally relevant to action *properties*. Failure to appreciate the difference between the two arguments can lead to inappropriate responses to the first argument. I show how Searle falls into this trap in some of his writing on "Intentional causation." Third, Searle's Chinese Room Argument continues to strike me as a sound argument. I give a formulation of it that makes certain tempting responses to it somewhat less appealing.

2. There is, I think, a *prima facie* tension between two things Searle says about consciousness and empirical theories of the brain. The tension is *prima facie,* as far as I am concerned, because it is not clear how exactly to understand what Searle says.

Comment on J. R. Searle (1993). The problem of consciousness. *Consciousness and Cognition* 2, 310–319.

1053-8100/93 $5.00

[1] An empirical theory of the brain should be able to explain several essential features of consciousness, and in particular it should be able to explain intentionality (''that feature of many of our mental states by which they are directed at, or about states of affairs in the world'' (Searle, 1993).

[2] The Connection Principle: it is conceptually necessary that in order for something ''to be genuinely an intentional state it must be accessible in principle to consciousness'' (Searle, 1993), or, in order for something to be genuinely an intentional state it must be able to cause ''that state in a conscious form'' or else be actually conscious (Searle, 1990, p. 634), or, ''all intentionality must be accessible in principle to consciousness'' (Searle, 1990, p. 585).

An empirical theory of the brain, for the purposes of the present discussion, is one that shows ''[h]ow exactly . . . neurobiological processes in the brain cause consciousness'' (Searle, 1993). Roughly put, the *prima facie* tension is that empirical explanation is governed by a contingency constraint (you cannot give empirical explanations of conceptual truths), but the relation between consciousness and intentionality is conceptually too tight to permit an empirical theory of the brain to explain intentionality.

Empirical theories are in the business of explaining features of the things in their domain. Consciousness is so peculiar, however, that it is utterly baffling how brains could cause consciousness, and hence it is utterly baffling how there could be an empirical theory of the brain that explained anything about consciousness. Searle's major contribution to the empirical and philosophical study of consciousness is to insist simultaneously on making clear just what is so baffling and on resisting the temptation to conclude either that there is no such thing as consciousness or that it is *so* peculiar that it will remain forever out of reach of scientific understanding.

Let me sketch part of what is so peculiar about consciousness, in order to provide a background for evaluating claim [1]. The crucial feature of consciousness that makes it so baffling is subjectivity. Why should subjectivity be so baffling? Very crudely put, empirical theories explain phenomena by relating those phenomena to other phenomena by laws of nature. To confirm an empirical theory, a theorist must be able to observe the phenomenon to be explained and the various phenomena by which it might be explained. But the theorist cannot observe the subjective character of conscious mental states. She cannot observe the subjective character of others' mental states, since it is essential to their subjectivity that they can be directly observed only by their owners. Nor can she observe the subjectivity of her own mental states, since ''any observation that [she] might care to make is itself that which was supposed to be observed'' (Searle, 1992, p. 99). Hence it seems utterly mysterious how the theorist could possibly confirm an empirical theory of consciousness.

An empirical theory of the brain that shows how brains cause consciousness must therefore be a very special sort of theory—but there is no reason to suppose

there could not be any such theory. Two things are needed. One is knowledge of empirical regularities between states of consciousness and states of the brain. We have a lot of knowledge of that sort, and we are gaining more every day. The other is some sort of conceptual breakthrough that will enable us to see how neurobiological states could cause conscious phenomena.

It remains difficult in our current conceptual bind to know quite what we are talking about when we advert to explanatory concepts we do not now possess, and kinds of explanations we are not now in a position to understand. But I think we can say a little about what any empirical explanation must be like, even if it is an explanation that uses concepts we do not now understand. An empirical explanation must rely on merely contingent connections between the phenomena that do the explaining and the phenomenon to be explained. For instance, an empirical explanation of blindsight might show how information available at sensory surfaces can influence guessing behavior even if no qualitative character comes along with the information. In this case, the connection is perfectly contingent. But consider by contrast the relation between being a liquid and flowing. Necessarily, if something is a liquid then, in suitable circumstances, it flows. Hence no empirical explanation could explain something's flowing on the basis of the fact that it is a liquid. For all that, there can be an empirical explanation of why something flows, and hence counts as a liquid, based on contingent facts. Water flows, and hence is a liquid, because aggregates of H_2O molecules have certain features that permit them to flow in suitable circumstances.

The Connection Principle, claim [2], states a conceptual connection between intentionality and consciousness. The argument for the claim (Searle, 1990) depends on an essential feature of intentional states. Each intentional state is directed at or about states of affairs, and each intentional state presents the state of affairs at which it is directed under some aspect or other. Any state of affairs in the world can be described in indefinitely many ways, but an intentional state presents it in exactly one way. Searle calls this essential feature of intentional states their "aspectual shape."

How do intentional states come by their unique aspectual shape? Consider unconscious intentional states, like your belief, most likely not conscious until you read the following, that Encino is west of Salt Lake City. There is nothing more to being such a state than facts available to any competent observer—what Searle calls "third-person" facts. But (Searle claims) no set of such facts can constitute the aspectual shape of the state. But unconscious mental states really do have an aspectual shape—how can this be? Clearly, conscious mental states have an aspectual shape, since when we are conscious of things we are conscious of them under one aspect and not another. We could explain how an unconscious intentional state has an aspectual shape if we maintain that the aspectual shape of an unconscious intentional state is the aspectual shape it has when it is conscious; but to make this claim we need to hold that unconscious intentional states are in principle accessible to consciousness. (The argument for why *only* this connection to consciousness can bestow a state with aspectual shape is enormously complex and mostly implicit in Searle's writing; it depends on showing in detail that attempts to show how to "constitute" aspectual shape from "third-

person facts'' all fail, and that the system of failures shows that no such attempt can succeed.)

Now, if one *does not* subscribe to the Connection Principle *at all* then there is no obstacle (yet mentioned) to an empirical theory of the brain and consciousness that also explains intentionality. Perhaps there are features of the mechanisms underlying consciousness that play some role in the mechanisms that enable one to have intentional states.

Searle, of course, cannot take this route. Nor can he claim that an empirical theory of the brain is at the same time an empirical theory of intentionality, if this is supposed to mean that there is, after all, a way of showing that aspectual shape can be constituted by third-person facts.

What else could he be asking of an empirical theory of the brain? Somehow, presumably using conceptual resources we do not now possess, the theory shows how brains cause consciousness. Most conscious states are intentional, and all intentional states are at least potentially conscious. So the idea would be to show, by looking at the contingent empirical relations by which brains cause consciousness, how brains cause intentionality.

The trouble is that the situation appears to be just the same as the situation with an empirical theory of liquidity. Liquidity is a molar feature of aggregates of various kinds of molecules. We can hope to give empirical explanations for how aggregates can have the properties needed in order to be liquids. But we cannot hope to get an empirical account of how the potential to flow hooks together with liquidity, since that is already guaranteed by the concepts. If Searle is right, consciousness and intentionality are molar features of certain kinds of physical objects. We could hope to give empirical explanations for how these objects can have the properties needed in order to cause minds. But we cannot hope to get an empirical account of how consciousness and intentionality hook together, since we already know that they are conceptually connected.

That is the *prima facie* tension. It may be that there is no real tension here at all. Perhaps Searle is simply claiming that there could be an empirical theory of the brain that, first, entails a large number of laws relating neurobiological phenomena and conscious phenomena, and second, shows how brains can have the structural features required in order to instantiate the (conceptually necessary) structural features of the ways consciousness and intentionality are related.

3. I have been discussing one version of a vexing problem: where exactly is the boundary between conceptual (or philosophical) inquiry and empirical inquiry? In this section I want to discuss another version of this problem and show how Searle's views on intentionality can be improved by careful attention to the boundary.

Consider the following argument about psychology and causation (I will call it the ''Original Logical Connections Argument''):

> Causation must be contingent; we learned that from Hume. But the connections among perception, thought and action are not contingent. For example, reasons for action are logically connected to the actions they explain, since what individuates a reason as the reason it is is the fact that it represents that action. Or, again, what it means to say that a

person has a certain reason is simply that the person would do certain things in certain circumstances. Therefore, reasons cannot be causes or effects.

Versions of this argument were popular with behaviorists and crypto-behaviorists (e.g., Melden, 1961). Philosophers and psychologists have long since recognized that this is an unsound argument (see Davidson, 1963, for the classic response to the argument). It trades on a confusion between events and descriptions of events. The fact that events have descriptions that are logically related has no bearing on whether those events are causally related. If it had any bearing, we could show that no events are causally related. Take any two events, *c* and *e*, and suppose that *c* causes *e*. *c* is correctly described as "the cause of *e*," and the sentence "*c* is the cause of *e*" logically entails that *e* occurred.

Although unsound, there was something right about the Original Logical Connections argument, and I want to try to articulate what it was. Its conclusion concerns causes; that is, it concerns particular events. There is another argument (I will call it the "New Logical Connections Argument"), quite similar to it and based on similar concerns, that concerns properties of events (Dardis, 1993, gives a more detailed presentation of the background to the argument). To describe the argument, I need to begin by making explicit an account of the metaphysics of ordinary causal explanations.

The world is full of changing things. Causal relations hold between changing things: changes in one thing make other things change. I toss a brick at a window and the window breaks. To explain what happens we need regular connections between changes in things. Why does the brick break the window? Well, other things equal, pretty much anything that applies at least a certain amount of force to a very small area of something as fragile as a pane of glass will cause the pane to shatter. So application of a certain force is regularly connected with the shattering of fragile things. Regular connections obtain between repeatable features of events. For convenience I shall call these repeatable features "properties." If we want to explain why the brick breaks the window, we appeal to some properties of the brick's being tossed rather than to others. Indefinitely many properties of the brick's being tossed through the window do not matter to the window's breaking at all. As it happens the brick I toss is made in Akron, Ohio, but there is no reliable connection between movements of things that were made in Akron, Ohio, and the breaking of windows. For terminological definiteness, let me call a property of a cause that matters to a property of the effect a "causally relevant property." Causal relevance should be thought of as a relation on pairs of properties.[1] Causal regularities obtain between certain pairs of properties and not others. In any causal transaction, only certain properties of the cause matter to the effect; equally, they matter only to certain properties of the effect. Just as it is the property of the brick that it had a certain momentum that mattered to the breaking of the window (rather than the fact that the brick was made in Akron), that property matters to certain properties of the effect (the fact that the

[1] Hence these two locutions are interchangeable: property P is causally relevant to property Q; the pair, <P, Q>, is in the causal relevance relation. No property is *simply* causally relevant.

window broke) rather than to others (the fact that something happened to the only way to see Main Street from inside the house).

What makes it the case that one property is causally relevant to another? The basic idea is the one that Mill's Methods aim to capture, as for instance the First Canon (Mill, 1872/1963, Book III, Chap. viii, Sect. 1):

> If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance, in which alone all the instances agree, is the cause (or effect) of the given phenomenon.

Mill's terminology betrays the familiar confusion between events and properties of events; once this is noted, the claim is that what makes one property causally relevant to another is the fact that what is common, in some appropriate sense, to all instances of some property of effects is some one property of the causes. Of course, much more needs to be said, and in fact I think most would agree that we do not at present have an acceptable fully articulated version of this basic idea (e.g., Cartwright, 1983; Eells, 1991).

All attempts to produce fully articulated versions of the basic idea need to include contingency requirements on the relation on properties. The reason for including such contingency requirements is that without them some properties would satisfy the account of causal relevance simply because they logically require that the relevant empirical pattern is satisfied by instances of those properties. The simplest case is the one alluded to above that shows that the Original Logical Connections Argument was unsound. Suppose c caused e. Then e has the property of being caused by c, and c has the property of being c.[2] But surely *being c* is not causally relevant to *being caused by c*. Or again, if something has the property of being sunburned then it is something with a distinctive sort of burn, and that burn must have been caused by exposure to the sun. But then the relation between being exposed to the sun and being sunburned is not contingent. Hence the propety of being an exposure to the sun is not causally relevant to the property of being sunburned. Although the property of being an exposure to the sun cannot be causally relevant to the property of being sunburn, it is causally relevant to another property that any instance of sunburn has: the property of being a certain sort of burn. (Actually, that will not quite do either, since being a burn is being a state that has a cause of a certain loosely specified sort, so the same problem arises again. But there must be some property of burns that is conceptually independent of any sort of cause.)

Of course, all these terms ("cause," "causal relation," "causal relevance") are terms of art and terms with a sordid history. So we cannot rely on intuitions about whether it would be appropriate to use various locutions in evaluating what I have just claimed. Exposure to the sun does, of course, cause sunburn. In *that* sense exposure to the sun is causally relevant to sunburn. But that is not the sense I intend the phrase "causal relevance" to have. The phrase "causal relevance" was introduced to denote a contingent relation on properties.

[2] I assume a liberal conception of properties, according to which pretty nearly any predicate, naturally occurring or constructed, from a natural language like English expresses a property.

As with any stipulative definition, it is important to defend the claim that the stipulation is not merely arbitrary. In this case the stipulation is intended to capture the metaphysical conception that lies behind Mill's Canons. The conception is that the natural world proceeds as it does because it is governed by natural causal laws. These laws relate repeatable properties of the things in the world. Their being *causal* laws amounts to their being only contingently true.

Here is an attempt at part of the contingency requirement on the causal relevance relation.

The Sunburn Condition

> If the property F is causally relevant to the property G, then G is "simply independent" of F; that is, there is a logically possible world in which there is a G-event that is not caused by an F-event.

The Sunburn Condition ensures that being exposed to the sun is not a candidate for being causally relevant to being sunburned. It permits being exposed to the sun to be causally relevant to one's skin being toasted.

The first attempt is not quite good enough. The properties *being exposed to the sun* and *being exposed to the sun for four hours* are different properties. And being sunburned is "simply independent" of being exposed to the sun for 4 hours, since you do not have to stay in the sun for exactly 4 hours to get sunburned. But if being exposed to the sun is not causally relevant to being sunburn, then neither is being exposed to the sun for 4 hours.

We need a stronger condition that is sensitive to the fact that one of these properties necessitates the other.[3] Here is my suggestion:

The Sunburn Condition

(Official Version)

> If the property F is causally relevant to the property G, then G is "necessitation-independent" of F; that is, there is a possible world in which there is a G-event that is neither caused by an F-event, nor by any event that is F′, for any F′ such that F necessitates F′.

Since the Sunburn Condition is closely related to principles on which the Original Logical Connections Argument depended, we should expect that it has some consequences for psychology. And it does. The causal theory of action holds that, necessarily, an event is an action only if it is caused by reasons.[4] If that theory is right, then the property of being a reason cannot be causally relevant to the property of being an action, since that pair of properties does not meet the Sunburn Condition. Moreover, being a reason with a particular content entails being a reason, so being a reason with a particular content cannot be causally relevant to the property of being an action, either. Hence the New Logical Connections Argument (which I hold to be sound):

[3] The property P necessitates the property Q just in case necessarily, if something has P then it has Q.

[4] The causal theory of action is, I take it, a piece of conceptual analysis, so if it is true it is necessarily true.

[3] The property of being an action is "necessitation-independent" neither of the property of being a reason, nor of any particular reason property. (The causal theory of action.)
[4] If the property F is causally relevant to the property G, then G is "necessitation-independent" of F. (The Sunburn Condition.)
[5] Therefore, neither the property of being a reason nor any particular reason-property is causally relevant to the property of being an action.

Even so, reasons cause actions. And the property of being a reason can still be causally relevant to non-intentional properties of actions.

Accepting the Sunburn Condition on causal relevance requires recognizing that being a causally relevant property and being a property suitable for mention in a causal explanation are two quite different things. There is a very indirect relation between causal relevance and causal explanations. Many causal explanations mention pairs of properties that are in the causal relevance relation: the momentum of the brick causally explains the shattering of the window. Some causal explanations, in particular, reasons-explanations of actions, depend for their force on the fact that some causally relevant properties are at work, but not all the causally relevant properties are explicitly mentioned by the explanations.[5] Take, for instance, this reasons-explanation: I threw the brick because I wanted to break the window. This is a genuinely causal explanation, since it is true only if my desire actually causes what I do. But I have just argued that the property, wanting to break the window, cannot be causally relevant to the property, throwing the brick, because there is a conceptual connection between the latter property and properties of the sort to which the former belong. How are causally relevant properties involved in this explanation? I think it is plausible to hold that the property, wanting to throw the brick, is a genuinely causally relevant property of the cause. The only plausible candidate for a property of the effect to which it is causally relevant seems to me to be the property of the effect, being a bodily movement of a certain (specific) sort.

If I am right, then, there are two distinct issues about the way causation and conceptual connections run together, one about particulars, one about properties, and there is some danger of not appreciating their distinction. And if one does not appreciate their distinction, then one may fail to address the issues in the right way.

Searle gives (Searle, 1979) a particularly useful and lucid description of the subject matter of any inquiry into intentionality. In the course of showing how not to make a variety of mistakes about intentionality, Searle comments on the

[5] Notice this distinction is orthogonal to another, related, distinction. We can distinguish *basic* and *non-basic* causal relevance. A pair of properties is in the basic causal relevance relation just in case the regularity linking them is basic to the causal structure of the world; a pair of properties is in the non-basic causal relevance relation just in case the regularity linking them is in some sense dependent on the basic regularities. Many ordinary causal explanations do not mention pairs of basic causally relevant properties. The point here is that many ordinary causal explanations do not mention pairs of properties that are in either causal relevance relation, since both kinds of causal relevance are governed by the Sunburn Condition.

Original Logical Connections Argument: the argument that claims "an Intentional state such as desire or wanting cannot be a cause of an action because desires and wants to do something are logically related to the doing of that thing, and hence, so the story goes, they cannot also be causally related." (p. 85) Searle's response goes like this:

> The way in which the desire to perform the action is logically related to the action is the way that any representation is logically related to the thing it represents. It is internally related to it in the sense that it couldn't be that representation if it didn't represent that thing. But this form of logical connection is no obstacle whatever to the connection also being a causal connection. Indeed, the way in which a desire to do something is causally connected to the doing of that thing is not only not inconsistent with there being the logical connection of representation, but it precisely requires that logical connection. It is not *in spite of* the fact that my desires are logically connected to my actions that my desires can cause actions; on the contrary, it's *only because* my desires are logically related to actions by way of representing them that they can be the sorts of causes of actions that they are [to wit, ones that explain actions in the special way that reasons explain actions]. . . . Wants motivate, but they motivate the things they are wants to do precisely because they are representations of those things, hence the causal connection requires a logical connection. (pp. 85–86, emphasis in the original)

Searle goes on to make two further points. (1) The situation with wants causing actions is exactly analogous to the situation with blueprints causing houses. For a blueprint to explain the construction of a house it must be the case that there is a "logical" or "internal" connection between the blueprint and the resulting house: the blueprint must be a representation of the house. But it would be mad to claim that blueprints cannot cause houses. (2) Some people, sensibly rejecting the arguments that reasons cannot be causes, hold that the way out of the apparent paradox (that something logically related to its effect can be causally related as well) is to hold that reasons and actions must have some features such that expressions that denote those features are not logically related. Searle holds that no such way out of the paradox is needed, since it is clear that there really is no paradox.

Let me consider these points in turn.

Of course, I agree that reasons are causes. But this diagnosis of what has gone wrong is, I think, somewhat misleading. There is an important disanalogy between reasons-explanations of actions and blueprint-explanations of houses. The crucial issue concerns the way in which wants explain actions. Actions must be both caused by and represented by the wants that are their motives. Hence necessarily, if something is an action, then it was caused by something that is a representation of it. It is not the case that necessarily, if something is a house then it was caused to exist in some way that involved a blueprint. Some houses are built with no blueprints; some houses (e.g., caves) are not built at all, but count as houses since they are usable as shelter. According to the Sunburn Condition, then, being a blueprint can be a causally relevant property with respect to being a house, while being a reason cannot be a causally relevant property with respect to being an action. Hence, while I think that everything Searle says here is true, it does miss a crucial point about the structure of the causal explanation of action.

On the second point, Searle writes (p. 86), "people who are aware that there is no logical obstacle to two things being both causally related and logically related believe that the way to remove the apparent paradox is to say that the terms of the logical relation have some more intrinsic characterization under which they can be characterized as in the causal relation." Searle dismisses the need for such a way out of the paradox. There are two different issues here; about one of them, Searle may be right, but about the other I think he is wrong.

What exactly is at issue between Searle and his opponent? Consider some causal explanation that denotes properties that violate the Sunburn Condition: "She's sunburned because she sat out in the sun for 4 hours." The "terms of the logical relation" in this case are two occurrences: her sitting out in the sun for 4 hours, and her getting sunburned. If these occurrences are ordinary particulars, that is, if they have indefinitely many properties (for instance, her sitting out in the sun for 4 hours occurred some precise number of kilometers from the Eiffel Tower), then clearly the two terms do have other characterizations than the ones offered in the explanation. So that cannot be the relevant issue.

The issue can be brought out by looking at another locution Searle offers on behalf of his opponent. Searle is here criticizing those who hold that "wants can figure as causes of actions *because* actions can be described in terms of their neurophysiology or some such" (p. 86, emphasis added). It is widely held that a full account of causation requires a specification of a distinguished set of ordered pairs of properties of events, such that when two events stand in the causal relation, what *makes it the case,* or *explains,* or *accounts for the fact* that they stand in the causal relation is the fact that they instantiate a pair of properties from this distinguished set (e.g., Johnston, 1985). Call pairs of properties from this set "causation grounding pairs." Searle argues (Searle, 1983, Chap. 4) against a particular version of this claim. A "neo-Humean" account of causation might hold that, necessarily, if two events are causally related, then they must be related by a strict causal law.[6] Searle's suggestion is that such an account is not warranted by the concept of causation. What makes it the case that one event causes another is the fact that the first makes the second happen; "making happen" need not be anything that requires the existence of a law.

This is not the place to attempt to evaluate Searle's suggestion. (In general it is hard to know what is and is not warranted by a concept. In this case the main question is whether we find it intelligible to suppose that one event could cause another, yet there be no more to this fact than that the first made the second happen. Many find this supposition unintelligible.) If he is right, then the point is fully general and has nothing to do with the special case of the Original Logical Connections Argument. But suppose he is wrong on the general point. The current context is a discussion of the Logical Connections arguments. His opponent

[6] Davidson (1967) urges exactly this "neo-Humean" claim. It is worth noting that Davidson does not explicitly endorse the claim that something "makes it the case that" two events stand in the causal relation. Hence it is unclear whether he intends this "neo-Humean" claim to be making the further claim that instantiation of a strict causal law "makes it the case that" two events are causally related.

should then be understood to be making a specific point that has particularly to do with the Logical Connections arguments. There is such a point, and I think it is a good one.

Suppose Searle's opponent is right that causation requires some such distinguished properties. The real issue now is whether "causation grounding pairs" could include a pair of properties, one of which cannot possibly be instantiated unless its instance is caused by an instance of the other. Put differently, consider a true explanation of an event, such that the explanation is causal, and moreover the descriptions used to pick out the events express properties that violate the Sunburn Condition. Could the fact that the two events stand in a causal relation be accounted for by the fact that they have the properties expressed by the descriptions used in the explanation, and moreover by no other fact? Could, for instance, the property of being a desire for a cold beer be paired with the property of being an act of reaching for a cold beer, such that the resulting pair of properties is a causation grounding pair?

Here is an argument that the answer is "no."

[6] Events are by nature independent: no event is such that its existence requires the existence of another event.

[7] Events have essential properties, and among them are the properties that ground the fact that they stand in causal relations (if they do).

[8] If there were a "causation grounding pair" which violated the Sunburn Condition, then the effect-event would have essentially a property that logically requires a cause of a certain sort, and hence the effect-event could not possibly occur without a cause. But that contradicts premise [6].

[9] Therefore, "causation grounding pairs" of properties must not include pairs which violate the Sunburn Condition.

If Searle's opponent is right that causation must be "grounded" and that "causation grounding pairs" must be governed by the Sunburn Condition, it follows that an action and the reason that causes it must have other characterizations under which they can be characterized as in the causal relation. I do not know whether Searle's opponent is right about these two assumptions, but they are plausible, and hence we cannot tell whether Searle's criticism hits the mark until we reach a decision on these basic metaphysical questions about causation.

4. Searle has long argued against the computational model of the mind, insisting that minds are not (just) programs and that brains really do matter to thought. I believe his Chinese Room Argument to this conclusion (Searle, 1980) is sound, and I want to offer a formulation of it that makes clear why I think it is sound.

Years ago students at the Bronx High School of Science learned to program on the school's IBM 1620. This machine had *real* core memory: it stored data in main memory by inducing changes in the magnetic fields of tiny iron doughnuts—the "cores"—strung on a grid of control wires. Consequently it emitted a good deal of radio-frequency radiation. Some clever students wrote a program that caused an AM radio perched on the front panel of the machine to play music.

Strong Artificial Intelligence (strong AI) is the claim that an appropriately pro-

grammed running digital computer literally thinks and that the program explains human cognition. Searle's Chinese Room Argument is specifically designed to show that strong AI is false. The following version of the argument allows that an appropriately programmed running digital computer could think, but that the program would matter to the thinking about as much as the program of the IBM 1620 matters to its ability to play music.

[10] Anything that has genuine intentional states (or is conscious) must have causal powers sufficient to produce genuine intentionality (or consciousness).

[11] Each Turing Machine is such that there is no causal power shared by all its realizations.

[12] Therefore, no Turing Machine is such that realizing it is sufficient for having genuine intentional states (or consciousness).

The argument for premise [10] is as follows (Searle, 1992, Chap. 4, especially p. 92). The contemporary scientific world view has it that the natural world is composed of physical matter and nothing else. That world view is not logically obligatory, but it is rationally overwhelmingly compelling in light of what we think we have learned about how the world works. Moreover, we (people with genuinely intentional states and genuinely conscious states) are part of the natural world. If this is the right view, then some very weak form of materialism about intentionality and consciousness must be true: intentionality and consciousness must have some causal basis in the causal order of the natural world.

Three comments about premise [10]:

a. One might worry that consciousness and intentionality might themselves be basic material phenomena, and hence that they could not be products of the causal powers of material phenomena. (Physics counts properties like charge as basic: those properties in terms of which all others are to be explained.) This is a legitimate worry, but one not relevant to the present argument, since the argument would remain valid even if we substitute for premise [10] the claim that intentionality and consciousness are basic material phenomena.

b. One might worry that the conclusion of the argument is secured by stealth, since it is entailed by the first premise alone. The problem has to do with the phrase "causal powers sufficient to produce." Premise [10] can be read to entail that intentionality and consciousness cannot be *functional* features of suitably organized aggregates of material stuff. But the computational theory of the mind is essentially the claim that intentionality and consciousness are functional features of things. To avoid the appearance of stealth I shall insist that premise [10] permits the possibility that the causal powers sufficient to produce genuine intentionality and consciousness are powers sufficient to produce a suitable functional organization.

c. Premise [10] (and the rest of the argument) is about both genuine intentionality and consciousness. The original Chinese Room Argument was couched only in terms of intentionality. This argument can be run for either intentionality or consciousness.

Premise [11] concerns Turing Machines. It is essential to the force of the argument that Turing Machines are abstract objects, for instance sets of tuples

of numbers. The argument for premise [11] is that any object whatsoever can realize any Turing Machine whatsoever. There are two ways to support this claim. The way Searle favors (Searle, 1993) is to note that what makes some particular physical object a realization of a Turing Machine is an assignment, by some observer, of an interpretation of the object, which maps its states and state-changes onto the states and state-changes of the Turing Machine. Clearly an observer can interpret any object in any way she likes so as to interpret it as a realization of any Turing Machine she likes. If this is the right way to support premise [11], there is a "deeper" objection to the computational theory of the mind: all computation is observer relative, and hence all attributions of computational states are "as if" attributions (there is no such thing as an intrinsic, observer-independent, computational state). Consequently any attributions of intentionality (or consciousness) to computational devices in virtue of their computational features must be "as if" attributions.

There is a weaker way to support premise [11] (see (Putnam, 1988, pp. 121–125) for an example of this strategy). It does not yield this "deeper" critique of computational theories of the mind, but it is just as effective. That way is simply to define "realization of a Turing Machine" in such a way that it turns out that every object is a realization of every Turing Machine.

Since every object is a realization of every Turing Machine, no realization of any particular Turing Machine is guaranteed to have any particular causal powers; hence if having some distinctive causal powers is necessary for having genuine intentional states (or consciousness) no realization of any particular Turing Machine is guaranteed to have those distinctive causal powers, and hence it is false that being a realization of any particular Turing Machine is sufficient for having genuine intentional states (or consciousness).

If this argument is correct, it would not show that computers could not think, or even that the Chinese Room does not understand Chinese. All it shows is that running a program, or instantiating a Turing Machine, is not by itself sufficient for genuine intentionality or consciousness. Equally, running the Bronx High School of Science music program is not sufficient for the music—running the same program on my up-to-date personal computer has no effect on the radio at all. The argument does show that what matters to the production of consciousness, or intentionality, is causal powers of the realization of the Turing Machine, rather than the fact that it is the realization of a Turing Machine.

One very natural response to this argument is to protest, "that couldn't be the correct way to understand what 'realization of a Turing Machine' means!" The idea would be to constrain realizations by introducing some abstract causal power that a realization of a Turing Machine must have. For instance, if the machine is supposed to go to state 917 when it is in state 23 and it receives an input of a '0', the input of the '0' is supposed to *cause* the machine to undergo this state transition, regardless of how the states and the inputs and outputs are realized.

This response concedes the objection to the computational theory of the mind made by my argument [10]–[12]. It says that something other than being a realization of a Turing Machine matters: not just any old thing can be a mind. Strong AI offers to explain human cognition in terms of programs that think. To constrain

what counts as a running program is to add physical features to the programs. If running such a program does explain cognition, then it must be that running the program is sufficient for cognition; but now the physical features added by the constraints are what explain cognition. It might even turn out (although Searle gives no argument that this is so) that realizations of Turing Machines have to be so tightly constrained that the ones that think can only be made out of biological stuff and one or two other things.

Premise [10] was stipulated to allow that the causal powers sufficient to produce genuine intentionality might be powers sufficient to produce a certain functional organization. I have just conceded that it might actually be correct that, given a suitable restriction on what counts as a realization of a Turing Machine, running the right program would be sufficient for thought and would explain human cognition. Hence I seem to be claiming that Searle believes that functionalism might be the correct solution to the mind/body problem. If anything counts as a *reductio ad absurdum* of my formulation of the Chinese Room Argument, that ought to. Well, if what I have conceded is functionalism, then so be it. Searle thinks it is an empirical question whether something made of silicon chips might have causal powers sufficient to produce genuine intentionality. If the answer is "yes," then it turns out that brains and certain aggregates of silicon chips share some nonbasic causal powers sufficient for genuine intentionality. That is functionalism, albeit a very anemic functionalism.

REFERENCES

Cartwright, N. (1983). Causal laws and effective strategies. In *How the laws of physics lie*. Oxford: Clarendon.

Dardis, A. (1993). Sunburn: Independence conditions on causal relevance. *Philosophy and Phenomenological Research,* **53.**

Davidson, D. (1980). *Essays on actions and events*. Oxford: Oxford Univ. Press.

Davidson, D. (1967/1980). Causal relations. *The Journal of Philosophy,* **64.** (Reprinted in Davidson, 1980.)

Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy,* **60.** (Reprinted in Davidson, 1980.)

Eells, E. (1991). *Probabilistic causality*. Cambridge: Cambridge Univ. Press.

Johnston, M. (1985). Why having a mind matters. In E. LePore & B. McLaughlin (Eds.), *Actions and events: Perspectives on the philosophy of Donald Davidson*. Oxford: Basil Blackwell.

Melden, A. I. (1961). *Free action*. London: Routledge & Kegan Paul.

Mill, J. S. (1872/1963). *Collected works VII: A system of logic*. Toronto: Univ. of Toronto Press.

Putnam, H. (1988). *Representation and reality*. Cambridge, MA: MIT Press.

Searle, J. R. (1979). What is an intentional state? *Mind,* **88,** 74–92.

Searle, J. R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences,* **3,** 417–457.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge Univ. Press.

Searle, J. R. (1990). Consciousness, explanatory inversion, and cognitive science. *The Behaviorial and Brain Sciences,* **13,** 585–642.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Searle, J. R. (1993). The problem of consciousness. *Consciousness and Cognition,* **2.**