

Can the mind change the world?

NED BLOCK

Hilary Putnam originated the idea that mental states are computational states. At first (Putnam, 1960), his view was that although mental states are not *identical* with computational states (or "logical states," as he then called them), there are useful analogies between them. Later (Putnam, 1967), he argued in favor of the identity on the grounds that it was more plausible to suppose mental states are functional states (as he then called them) than that they are behavioral or physical states. This doctrine – functionalism – has dominated the philosophy of mind for over twenty years. Shortly after proposing functionalism, Putnam rejected it again (1973), and he has maintained this position ever since (Putnam, 1988).

Putnam was my teacher during both my undergraduate and graduate days, and I fear I have absorbed his ambivalence toward functionalism. My teacher has had a habit of changing his mind, but never has he done so within a single essay, and so in this chapter I have surpassed him. My chapter starts out as an argument for functionalism, but it ends up suggesting an argument against it. The issue is whether we can avoid epiphenomenalism, which I here understand as the doctrine that what we think or want has no causal relevance to what we do. I propose functionalism as a way of warding off arguments for epiphenomenalism, but then I argue that functionalism may bring epiphenomenalism in its wake.

The orientation of the chapter is toward the sciences of the mind, and their relation to intentional content, that is, what is shared by the

I am grateful for support to the American Council of Learned Societies and to the National Science Foundation, grant number DIR88 12559. I am also grateful to Jerry Fodor, Paul Horwich, Frank Jackson, Gabriel Segal, and Marianne Talbot for their comments on an earlier draft; to Thomas Nagel, who was the commentator on the paper on which this chapter is based at a conference at Columbia University in December 1988; to Simon Blackburn, who was the commentator on the paper at Oxford in February 1989; to Martin Davies, Dorothy Edgington, and Barry Smith for discussion of the issues while I was at Birkbeck College, University of London; and to the audiences at a number of philosophy colloquia.

belief that grass grows and the desire that grass grows, the *that grass grows* that both states are directed toward. The question at hand is whether the sciences of the mind preclude intentional content from causal relevance to behavior. One argument that the intentional contents of our beliefs, thoughts, and the like have no effects on our behavior could be put this way: The processors in the head are not sensitive to content, so how could content have any effect on the outputs or changes of state of the system of processors? And if content can't affect the operation of this system of processors, how could it play any role in producing behavior? This argument seems formidable whether one thinks of the processors as neural devices reacting to neural inputs or, instead, from the cognitive science point of view, as computational devices processing representations.¹ In this chapter, I confine myself to the problem as it arises in the cognitive science approach that is dominated by the computer model of the mind. I assume a very specific picture of cognitive science and its relation to the commonsense conception of intentional content, namely, the view according to which there is an internal system of representation from whose meanings our intentional contents derive (Fodor, 1975; Pylyshyn, 1984). One of my reasons for couching the discussion in terms of this view is that although those who adopt this view are *motivated* by the aim of showing how our commonsense beliefs about content (including our belief in content's causal efficacy) are vindicated by the computer model of the mind, the problem of the epiphenomenalism of content arises within this view in an extremely simple and straightforward (and poignant) way. The viewpoint assumed throughout the chapter is that of a supporter of the computer model in cognitive science who also would *like* to believe that the contents of our thoughts are indeed causally relevant to what we do.

The problem I have in mind might be put in terms of *The Paradox of the Causal Efficacy of Content*, namely, that the following claims all seem to be true, yet incompatible:

- I. The intentional content of a thought (or other intentional state) is causally relevant to its behavioral (and other) effects.
- II. Intentional content reduces to meanings of internal representations.
- III. Internal processors are sensitive to the "syntactic forms" of internal representations, not their meanings.

The first claim is meant to be part of the commonsense view of the mind. The third is plausibly taken to be a basic claim of the computer model of the mind, and the second is a useful and plausible way of thinking how commonsense psychology meshes with the computer

model. I
question
useful p
I think th
premise

The r
Turing r
such as c
sensitive
But if th
a gate, h
gates? S
sentatio
influenc
soning a
compute
mind (c
assumin

My p!

I. F

II. E

ic

III. S

c

r

IV. I

g

A subt
relevan
of a law
a count

The fir
propert
some a
and ge
name :
Bureau
United
conten

e that grass at hand is intent from intentional contexts on our head are not effect on the d if content could it play formidable reacting to of view, as ; chapter, I tive science the mind. I lation to the he view ac- tation from '5; Pylyshyn, erms of this ated by the it (including ie computer i of content orward (and ter is that of o also would eed causally

e Paradox of g claims all

itional state) cts.
l representa-
c forms" of

view of the he computer sible way of ie computer

model. This second claim is by far the most controversial, but I won't be questioning it here. My reasons are that I think it is true, that I see no useful purpose to dividing meaning and content in this context, and that I think the best bets for resolving the paradox are to question the third premise and whether the reasoning that leads to the paradox is right.

The reasoning behind the paradox goes something like this: Any Turing machine can be constructed from simple primitive processors such as *and* gates, *or* gates, and the like. (See Minsky, 1967.) Gates are sensitive to the syntactic forms of representations, not their meanings. But if the meaning of a representation cannot influence the behavior of a gate, how could it influence the behavior of a computer – a system of gates? Since intentional content reduces to meanings of internal representations, and since meanings of internal representations cannot influence behavior, content cannot influence behavior either. The reasoning assumes that at least as far as our thinking is concerned, we *are* computers. This idea – which is simply the computer model of the mind (our cognitive mind, that is) – may be wrong, but I will be assuming it to explore where it leads.

My plan for the chapter involves

- I. Explaining each premise
- II. Examining and rejecting a putative solution based on a nomological conception of causal relevance
- III. Suggesting a solution based on a functionalist conception of content and meaning and a counterfactual theory of causal relevance
- IV. Discussing a problem with the proposed solution, one that suggests that functionalism actually breeds epiphenomenalism

A subtheme of the chapter is that a nomological theory of causal relevance (a theory that explains causal relevance in terms of the notion of a law of nature) has more of a problem with epiphenomenalism than a counterfactual approach.

1. The premises

The first premise uses the notion of a *causally relevant property*. Some properties of a cause are relevant to the production of an effect, and some are not. Hurricane Eliza broke my window. Eliza's wind speed and geographical path are causally relevant to the breaking, but its name and the location of its records in the United States Weather Bureau are not. According to the first premise, if my belief that the United States is a dangerous place causes me to leave the country, the content of the belief is causally relevant to the behavior; a property that

is not causally relevant to the behavior is the last letter of the name of the city in which the belief was formed.

Note that the point is *not* that beliefs, thoughts, desires, and the like (mental states or events that have content) are *causes*, for example, of behavior. (I assume that they are.) Rather, the point is that when mental events have effects, they typically have those effects (rather than different effects) *because* the mental events have the contents that they have, rather than some other contents. Typically, if the beliefs, thoughts, and so forth, had had contents that were appropriately different, they would have had quite different effects. For example, had I believed that everywhere except the United States is a dangerous place, then I wouldn't have left the country.

My metaphysical stance in this chapter is one in which mental events are the causes of behavior, and their contents are properties of those events that may or may not be causally relevant to the events' effects. I shall say that property *P* of event *c* is causally relevant to effect *e* and that *c* causes *e* in virtue of *P* more or less interchangeably. Also, I shall put the claim that content is causally relevant to something by saying that content is causally efficacious, and not epiphenomenal. Of course, my use of 'epiphenomenalism' is importantly different from the traditional one. I'm not raising any possibility of content being a property of a distinct mental substance, events in which are caused by events in the brain even though events in the mental substance never cause anything. And epiphenomenalism in my sense does not entail that content is itself caused by an underlying physical state that also causes behavior. The reason I use this old word ('epiphenomenalism') for this newer problem is that the problem I raise is the modern heir of the old problem.

The second premise – that content reduces to meanings of internal representations (symbols in the head) – is much more tendentious than the first, and is certainly not part of commonsense wisdom about the mind. But it is a straightforward way of making commonsense realism about content compatible with the view that the machinery of the mind is one of computations on internal representations. According to the computer model, the mind (or its cognitive aspect) can be thought of as a system of processors that take representations as inputs, transform them in various ways, and then send them to other processors, as in computers. One can think of the representations in such a system as being in certain computational relations to the whole system. These computational relations are determined by the ways the system would treat the representation given various different states of its component processors. A sample computational relation is that of storing a representation.

Of course
remember
computer
bering mi
that it can
storing a
doctrine
premise.
thought t
internal i
propositio
thinking t
the head.

Talk of
talk. The
situation,
the system
differs fro
tionalisms
controvers
tional.

The thi
syntactic
their mea
and gate i
inputs yie

- I. Is i
- II. Do
- III. Do

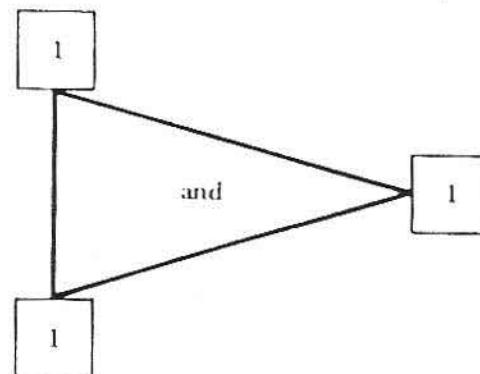


Figure 1

Of course, computers that you can buy don't currently think or remember. But if computers can be programmed to think, or if we are computers, this story can be extended to thinking things. Thus, remembering might consist in the storing of a representation in such a way that it can later be accessed. Remembering that grass grows would be storing a representation that means that grass grows. This is the doctrine of reduction of content to meaning stated in the second premise. More generally, as is argued in Fodor (1975), to have the thought that p is to be in a certain computational relation to an internal representation that means that p , and likewise for other propositional attitudes. The slogan – and it is only a slogan – is that thinking that grass grows is having 'Grass grows' in the thought box in the head.

Talk of the thought box in the head is (thinly) disguised *functional* talk. The sentences that are in the thought box share a computational situation, a role within the system – in other words, a function within the system. This functional theory of what a thought *is* (and how it differs from a desire) should be firmly distinguished from other functionalisms to be mentioned in this paper, especially the much more controversial idea that the meaning of a representation is itself functional.

The third premise is that internal processors are sensitive to the syntactic form of the internal representations that they process, not their meanings. Consider the *and* gate of Figure 1. What makes it an *and* gate is that it emits a '1' if and only if both inputs are '1's; all other inputs yield a '0' as output. The *and* gate

- I. Is not sensitive to
- II. Does not react to
- III. Does not detect

whether the '1's' represent truth or the number 1 or nothing at all. Rather, the *and* gate is sensitive to, reacts to, and detects only whether the inputs are both '1's' or not. Thus it is sensitive to the syntax, not the meaning of its inputs, and likewise for the primitive processors postulated by cognitive science accounts of the mind. (The primitive processors of a system are the ones for which there is no explanation *within* cognitive science of how they work; their operation can be explained only in terms of a lower level branch of science, physiology, in the case of human primitive processors, electronics in the case of standard computer primitives.)

Note that the sense of 'syntax' I am using here (somewhat misleadingly) means *form class*. '1' and '0' are different syntactic objects in this sense. It is important to be aware that syntax in this sense of the term is another functional notion. English orthography is also functional, although this may be obscured by the rigidification of function by convention. For %exampl%, you will hav% littl% troubl% figuring out what 1%tt%r of th% alphab%t is th% on% to which th% unusual symbol in this s%nt%nc% should b% tr%at%d as functionally %equival%nt.

Consider an input-output system whose input and output registers are bi-stable, and take on values of either 7 volts or 4 volts. Suppose that if both input registers are at 4 volts, then the output is 4 volts, and every other input yields the 7-volt output. Then (1) the system is an *and* gate, (2) the 4-volt value counts as a '1' for this *and* gate, and the 7-volt value counts as a '0'. A differently constructed *and* gate might be one for which 7 volts counts as a '1'. '1' is conventionally assigned to states of computer registers using this type of consideration. The functional roles of the bi-stable states of registers simultaneously determine our identifications of the devices in the system (e.g., as adders and gates) and our identifications of the states of the registers as symbols. So it is having a certain functional role that makes a state satisfy a *syntactic* description, in the sense of syntax used here.

Note that it would be a mistake to say that 4 volts in the first gate mentioned has the same *meaning* as 7-volts in the second gate. We don't know *what* the meanings of the '1's' in either gate are until we see other aspects of their function. These '1's' could be used to mean one, or true, or green; the input-output function does not choose among these and other possibilities. The aspects of function relevant to syntax are different from, though overlapping with, the aspects of function relevant to meaning.

In the next section, I will explain the picture common in cognitive science of sensible or "rational" relations among contents deriving from the correlation between rational relations among contents on the one

hand, an
the other
cognitive
potential
shall see
"going p
worries :
section is
to syntax
solution.

2.

The title
example,
handle o
only knov
represent
decimal
following

0 + 0
1 + 0
0 + 1
1 + 1

The fir
last is tru
Here is
'0' in this
gate. The
inputs ar
illustrated

hing at all.
only whether
tax, not the
isors postu-
tive proces-
ation within
e explained
in the case
of standard

at misleads
jects in this
the term is
ctional, al-
by conven-
ing out what
l symbol in
al%nt.

ut registers
ts. Suppose
4 volts, and
m is an *and*
d the 7-volt
ight be one
ed to states
functional
termine our
(and gates)
ols. So it is
a *syntactic*

he first gate
d gate. We
are until we
sed to mean
not choose
relevant to
aspects of

in cognitive
eriving from
on the one

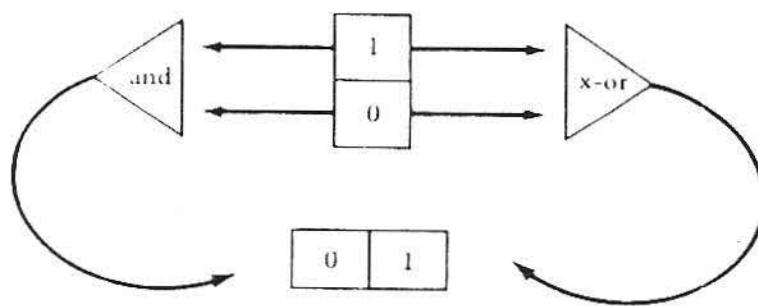


Figure 2

hand, and processing relations among syntactic objects in the brain on the other. This idea will illustrate both the strength and weakness of the cognitive science picture I will be assuming. The strength lies in its potential for a mechanical account of thought and reason, but as we shall see, it does so in a way that relegates syntactic properties to "going proxy" for semantic properties, and that opens it up to the worries about epiphenomenalism discussed here. My aim in the next section is to motivate premise 3 – that internal processors are sensitive to syntax, not meaning – since it will take the heat in my proposed solution.

2. The brain as a syntactic engine driving a semantic engine

The title you just read can be understood by attention to a simple example, a common type of computer adder stripped down so as to handle only one-digit addenda. To understand the example, you need only know the following simple facts about binary notation: 0 and 1 are represented alike in binary and decimal, but the binary translation of decimal '2' is '10'. The adder pictured in Figure 2 will solve the following four problems:

$$\begin{aligned} 0 + 0 &= 0 \\ 1 + 0 &= 1 \\ 0 + 1 &= 1 \\ 1 + 1 &= 10 \end{aligned}$$

The first three equations are true in both binary and decimal, but the last is true only in binary.

Here is how the adder works. The two digits to be added, a '1' and a '0' in this case, are connected both to an *and* gate and an *exclusive-or* gate. The latter gate is a "difference detector," i.e., it outputs a '1' if its inputs are different, and a '0' if they are the same. In the case illustrated in Figure 2, the *exclusive-or* gate sees a difference, and so it

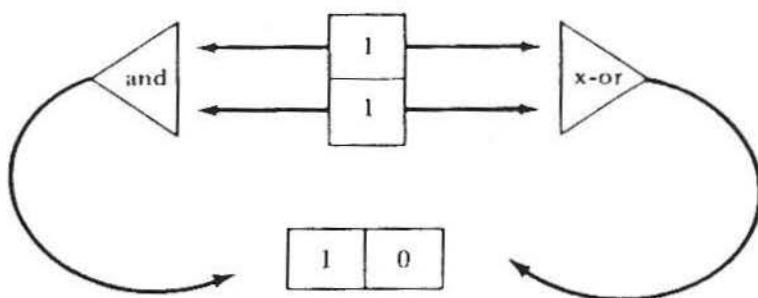


Figure 3

outputs a '1' to the rightmost box of the answer register. The *and* gate outputs a '0', and so the device computes the answer. The *exclusive-or* gate does the "work" in the first three problems, and is needed only for carrying, which comes in only in the last problem as illustrated in Figure 3. Both inputs are '1's, and so the *and* gate outputs a '1' to the leftmost box of the answer register. The other gate makes the rightmost box a '0', and so we have the answer.

Seeing the adder as a syntactic engine driving a semantic engine requires noting two functions: one maps numbers onto other numbers, and the other maps symbols onto other symbols. The latter function is concerned with the numerals as symbols – without attention to their meanings. Here is the symbol function:

$$\begin{aligned} '0', '0' &\rightarrow '0' \\ '0', '1' &\rightarrow '1' \\ '1', '0' &\rightarrow '1' \\ '1', '1' &\rightarrow '10' \end{aligned}$$

This symbol function is mirrored by a function that maps the numbers represented by the numerals on the left onto the numbers represented by the numerals on the right. This function will thus map numbers onto numbers. We can speak of this function that maps numbers onto numbers as the *semantic* function, since it is concerned with the meanings of the symbols, not the symbols themselves. (It is important not to confuse the notion of a semantic function in this sense with a function that maps symbols onto what they refer to; I shall discuss this latter function shortly.) Here is the semantic function (in decimal notation – you must choose some notation to express any function):

$$\begin{aligned} 0, 0 &\rightarrow 0 \\ 0, 1 &\rightarrow 1 \\ 1, 0 &\rightarrow 1 \\ 1, 1 &\rightarrow 2 \end{aligned}$$

The first
maps the
onto the

The ke
two funct

I. A

II. Sy

III. Us

The symt
meanings
useful re
function
described
these two
usual sen
how it is
numbers.
semantic
of symbol
The idea
nature ha
tions amc
of the sy
"syntactic
and 1's t
meaning-
– proces
a correlat
their fort
drive our

Now t
driving a
intention
our cogni
the synta
causal wo

In this s
making tl

The first function maps symbols onto symbols; the second function maps the numbers referred to by the arguments of the first function onto the numbers referred to by the values of the first function.

The key idea behind the adder is that of a correlation between these two functions. The designer has joined together

- I. A meaningful notation (binary notation)
- II. Symbolic manipulations in that notation
- III. Useful relations among the meanings of the symbols

The symbolic manipulations correspond to useful relations among the meanings of the symbols – namely, the relations of addition. The useful relations among the meanings are captured by the semantic function above, and the corresponding symbolic relations are the ones described in the symbolic function above. It is the *correlation between these two functions* (which establishes a semantic function in the more usual sense of a function from words to their referents) that explains how it is that a device that manipulates symbols manages to add numbers. Now the idea of the brain as a syntactic engine driving a semantic engine is just a generalization of this picture to a wider class of symbolic activities, namely, the symbolic activities of human thought. The idea is that we have symbolic structures in our brains, and that nature has seen to it that there are correlations between causal interactions among these structures and sensible relations among the meanings of the symbolic structures. The primitive processors “know” only the “syntactic” form of the symbols they process (e.g., what strings of 0’s and 1’s they see), and not what the symbols mean. Nonetheless, these meaning-blind primitive processors control processes that “make sense” – processes of decision, problem solving, and the like. In short, there is a correlation between the meanings of our internal representations and their forms. And this explains how it is that our syntactic engine can drive our semantic engine.²

Now the picture just sketched of the brain as a syntactic engine driving a semantic engine reveals how it is that a mechanistic theory of intentionality can invite the charge of epiphenomenalism. It seems that our cognitive processes exploit a *correlation* between the semantic and the syntactic. The syntactic properties of the representations do the causal work, and the semantic properties come along for the ride.

3. The appeal to laws

In this section, I will examine a putative solution, that is, a way of making the cognitive science picture (premises 2 and 3) compatible with

the causal relevance of content (to the behavioral and other effects of contentful mental states, of course – I'll be leaving the prepositional phrase out often, just speaking, elliptically, of the causal relevance of content). Actually, I shall start by briefly mentioning a reductionist proposal just to set it to one side. If content properties could be identified with, say, neurophysiological properties, then there would be no opening for epiphenomenalism. If content properties are simply identical to neurophysiological properties, then the causal efficacy of the neural would guarantee the causal efficacy of content. Whatever the merits of physiological reductionism, it is not available to the cognitive science point of view assumed here. According to cognitive science, the essence of the mental is computational, and any computational state is "multiply realizable" by physiological or electronic states that are not identical with one another, and so content cannot be identified with any one of them.³

Note that in rejecting this putative solution, I am not rejecting a "physicalistic" point of view. If all the nomologically possible things that can have computational properties are physical things, then the computational point of view, embracing this idea, is itself physicalistic. Even if the computational properties that characterize mentality are in this sense *physical*, that does not make them *physiological or electronic* (or syntactic, for that matter), and so physicalism in this sense does not lead to any suggestion that processors in the head can detect content or meaning.

There is another putative solution to which I will devote more attention, one that appeals to a nomological view of causation. The idea is that there are non-strict psychological laws involving content, and that law – even non-strict law – makes for causal relevance. Fodor (1987b, forthcoming) argues that intentional laws provide non-strict intentional sufficient conditions for behavior, and that is what makes content causally relevant to behavior. It will pay us to examine a simple version of this nomist perspective: *F* is causally relevant to an effect *e* if the instantiation of *F* is nomologically sufficient for *e* – even if the nomological sufficiency holds only *ceteris paribus*.

The trouble with this simple version of the nomist approach is familiar: there can be correlation without causation – even nomological correlation of *F* with *G* without a causal relevance relation between *F* and *G*. And nomological correlation can involve nomological sufficiency. Suppose *C* (for cause) is nomologically sufficient for and uncontroversially causally relevant to *E* (for effect). Suppose *X* is nomologically correlated with *C* because *X* and *C* share a causal source, and so *X* is not causally relevant to *C*. Then, *X* is nomologically sufficient for *E*

without I
"fork" c;

Here i
fire to a
enough
explosio
enough t
will expl

Now t
and elec
electron
electrica
cient for
conducti

I take
notion o
the nom
crease d
explosio
rising ele
relevant
that allo
to explo

The o
avoid th
cases ha
cases. ai
terms of
way to g
about w
be one t

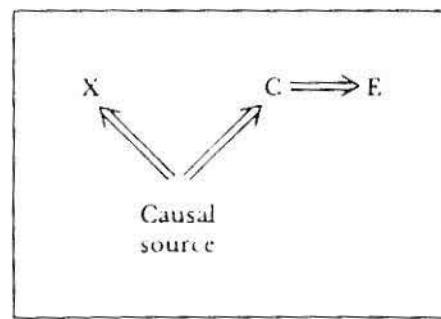


Figure 4

without being causally relevant to E. See Figure 4, which pictures this "fork" case.

Here is an example. Consider a setup in which a metal rod connects a fire to a bomb. So long as the thermal conductivity of the rod is low, not enough heat is transferred from the fire to the bomb to cause an explosion. But if the thermal conductivity of the rod is increased enough (say, by altering its composition), then the heat from the fire will explode the bomb.⁴

Now there is a law – the Wiedemann-Franz Law – linking thermal and electrical conductivity under normal conditions. (The same free electrons carry both charge and heat.) Hence for this setup, rising electrical conductivity, together with other things being equal, is sufficient for an explosion. Hence by the nomist criterion, rising electrical conductivity is causally relevant to the explosion.

I take it that to the extent that we have a pretheoretical grasp on the notion of causal relevance, this consequence of the simple version of the nomist criterion is simply wrong. The electrical conductivity increase does not cause the explosion; certainly it does not cause the explosion *in virtue of being a rise in electrical conductivity*. Rather, the rising electrical conductivity is an inactive concomitant of the causally relevant rising thermal conductivity. It is the rising thermal conductivity that allows more heat to be conducted to the bomb, causing the bomb to explode.

The obvious nomist response would be to reformulate the thesis to avoid these fork cases. The idea would be to formulate what the fork cases have in common that distinguishes them from genuine causal cases, and use this result to produce a definition of causal relevance in terms of "non-fork" nomological sufficiency. I think this is a promising way to go about characterizing causal relevance, but I have two doubts about whether the resulting nomist conception of causal relevance will be one that saves us from epiphenomenalism.

First, recall that the problem of epiphenomenalism arose to begin with because the cognitive science point of view held that nature had produced a correlation between syntax and semantics, only the former being directly causally related to behavior. What reason is there to think that an independently motivated definition of "non-fork" nomological sufficiency would count semantics as any more causally relevant to behavior than rising electrical conductivity is to the explosion in the example considered earlier? Indeed, the correlation between semantics and syntax via a common cause is a good example of a fork.

Second, -the issue of the epiphenomenalism of the semantic may simply resurface in a different form in an independently motivated account of causal relevance in terms of nomologicity. Horwich (1987, Chapter 8) frames a definition of causation in terms of the idea of a basic law. A cause is linked to its effects by chains of *direct* nomological determination, and direct nomological determination is determination via basic law. "A direct cause of some effect is an essential part of an antecedent condition whose intrinsic description entails, via basic laws of nature, that the effect will occur" (*ibid.*). Cause c causes effect e if there is a set of events, $e_1 \dots e_n$ linked by basic laws, and if $c = e_1$, and $e = e_n$. The account concerns event causation, not the causal relevance of properties. But assuming that it could be supplemented to form an account of causal relevance using the notion of a basic law, we can see that although it shifts the terms of the discussion of epiphenomenalism, it does not solve the problem. Is content epiphenomenal? That depends on whether the laws that determine behavior in terms of content are basic. But one's decisions on whether these laws are basic – and, indeed, what basicness *is* – will be (and should be) conditioned by one's decisions on what is causally relevant to what.

Of course the fact that the nomist conception of causal relevance will not solve this problem all by itself is not a reason to reject the nomist conception.

Fodor (forthcoming) says that if there is a *causal* law that F instantiations cause G instantiations (*ceteris paribus*), then F is causally relevant to the G instantiations that are caused by F instantiations. The appeal to causation in this sufficient condition is harmless for Fodor's purposes, but not for mine. For how are we supposed to know whether the laws that relate content to behavior are *causal* laws? For example, consider the (very low quality) law to the effect that instantiations of wanting G and believing that A is required for G cause instantiations of A, *ceteris paribus*. To know whether the content properties are causally relevant to the action, we must know if the law is causal. But if Fodor's sufficient condition is all we have to work with, the issue of whether the

law is causa
no progress

We have jus
of nomologi
nalism. I kno
relevance. n
section, I wi
– the Coun
vance is con
way of avoid

Suppose v
a {1, '0} pa
a '0' output,
just the sam
and '0' had
these symbo
been using
primitive pr
object (say, a
another strir
caused the s:
philosophica
the meaning
conclusion is
nal, so are th

I want to c
tions are the
for epiphenc

The functi
thinking, pla
bines with a
between inp
arise from th
functionalists
body. others
these imping
ists take fun
partly intern

arose to begin
that nature had
only the former
on is there to
"n-fork" nomo-
usally relevant
explosion in the
reen semantics
ork.

semantic may
tly motivated
orwich (1987,
the idea of a
-t nomological
determination
tial part of an
via basic laws
ses effect e if
if $c = e_1$, and
usal relevance
ed to form an
w, we can see
enomenalism.
That depends
f content are
basic – and,
conditioned by

relevance will
et the nomist

at F instantia-
isally relevant
s. The appeal
Fodor's pur-
whether the
For example,
stantiations of
stantiations of
s are causally
But if Fodor's
f whether the

law is causal just *is* the epiphenomenalism issue we are discussing, so no progress has been made.

4. A proposal

We have just been considering an approach to causal relevance in terms of nomologicity, and we have not succeeded in avoiding epiphenomenalism. I know of only one other remotely promising approach to causal relevance, namely, the approach in terms of counterfactuals. In this section, I will first give a version of the argument for epiphenomenalism – the Counterfactual Argument I will call it – in which causal relevance is construed counterfactually. Then I will suggest a functionalist way of avoiding the conclusion.

Suppose we have a computational device (say, an *and* gate) to which a {‘1’, ‘0’} pair – representing the numbers 1 and 0 – is input, yielding a ‘0’ output, representing 0. Now the {‘1’, ‘0’} input pair would have had just the same effect, viz., the production of ‘0’ as output, even if the ‘1’ and ‘0’ had represented truth and falsity instead of 1 and 0, or even if these symbols had represented black and white, or even if we hadn’t been using them to represent anything at all. And what holds for a primitive processor also holds for any system of them. If a syntactic object (say, a string of ‘0’s and ‘1’s) is input to the whole system, and another string of zeroes and ones is output, the input string would have caused the same output string no matter what it had meant. In common philosophical parlance, the syntax of the representation “screens off” the meaning from having any causal relevance to the output. The conclusion is that since meanings of representations are epiphenomenal, so are the contents of the mental states that they ground.

I want to offer the claim that if the meanings of internal representations are their “functional roles,” then the Counterfactual Argument for epiphenomenalism can be vanquished.

The functional role of a representation is its causal role in reasoning, thinking, planning, and in general in the way the representation combines with and interacts with other representations so as to mediate between inputs and outputs. The functional roles of representations arise from the ways processors manipulate their syntactic forms. Some functionalists take inputs to be impingements on the surfaces of the body, others take inputs to be the things in the world that produce these impingements (and likewise for outputs). Thus some functionalists take functional role to be internal, whereas others take it to be partly internal, partly external. My use of the functionalist perspective

requires no commitment on this matter. More specifically, nothing I say here should be construed as an endorsement of "narrow content."

A functional role theory of meaning is a theory of what meaning *is* that yields an account of what it is about representations that gives them their meanings, not a "semantics" in the sense of a theory of particular constructions in particular languages. Thus, it would not be the job of a functional role theory of meaning to explain why it is that 'The temperature is 70°', and 'The temperature is rising' do not entail '70° is rising'. Theories of the sort that would deal with such questions often address themselves to what meaning is, but they do not yield an account of what gives representations their meanings. Usually, for their purposes any "surrogate" of meaning – anything that acts like meaning in the way that set theoretic entities "act like" numbers – will do. (See Block, 1986, 1987, for more on this distinction.) A functional role theory can be thought of as a "use" theory of meaning, where the uses are at least partly inside the head.

One motivation for such an account can perhaps be seen more clearly if one reflects on a case in which one learns the meanings of words without anything approximating an eliminative definition for them. Consider, for example, the learning of a new scientific theory with its new theoretical terms, e.g., 'force', 'mass', 'momentum', 'energy'. These new terms are not definable in everyday language (or anything like an "observation language"), though they are definable in terms of one another. The student learns these terms by coming to understand how to use them in thought, in experiment and observation, and in solving problems on quizzes. If meaning is functional role, then it is easy to see why learning new terms is acquiring their use. It is a plus for a theory of what meaning is if it also tells us what it is to know and learn meanings.

That is all I will say here to motivate the functional role account.⁵ Now on to the Counterfactual Argument.

You will recall that the Counterfactual Argument said that it appeared that a given syntactic object would have caused the same output even if it had meant something quite different from what it actually means or even if it had meant nothing at all, so the syntactic form of a representation screens off its meaning from having any causal relevance.⁶

Suppose, for example, that the sentence 'There is danger coming this way' is in the "belief box" (with its normal meaning), causing one to flee. According to the Counterfactual Argument, the sentence would have caused the fleeing even if it had meant that my long-lost friend approaches, or that Empedocles leaped. But if functional role semantics is correct, then it is not at all guaranteed that 'There is danger

coming thi
Empedoc
meaning a
functional
functional

Perhaps
includes a
here wher
hence its
causal cha
coming thi
friend app
fleeing wo
avoidance,
sentence i
been differ

In sum,
tation wou
been differ
different fu
Counterfa

In terms
can be sen
Are the m
inputs?

Of cours
Thinking c
processor
representa
involve dif
representa
would have
meaning, t
for a gen
processor s
complex p
would have
meaning. I
possibility

This poi
tonomous
1980). Obs

lly, nothing I say
w content."
what meaning is
tions that gives
e of a theory of
it would not be
in why it is that
ng' do not entail
h such questions
do not yield an
Usually, for their
acts like mean-
bers - will do.
A functional role
, where the uses

een more clearly
anings of words
ition for them.
c theory with its
, 'energy'. These
anything like an
in terms of one
understand how
i, and in solving
it is easy to see
is for a theory of
learn meanings.
il role account.⁵

said that it ap-
the same output
what it actually
itactic form of a
any causal rele-

nger coming this
, causing one to
sentence would
long-lost friend
onal role seman-
There is danger

coming this way' would have caused fleeing even if it had meant that Empedocles leaped. For if it had had a different meaning, or no meaning at all, its functional role would have been different, and since functional role is causal role, abstractly construed, a difference in functional role typically will include a difference in behavioral effects.

Perhaps the functional role of 'There is danger coming this way' includes an inference to sentences such as 'It would be best not to be here when the danger arrives' (with its normal functional role and hence its normal meaning), and perhaps this inference is part of the causal chain that led to fleeing. Then if the token of 'There is danger coming this way' that was in the belief box had meant that a long-lost friend approaches, the inference would not have taken place, and the fleeing would not have occurred. We would have approach instead of avoidance, and friendly words instead of fear. If the meaning of the sentence in the belief box had been different, its effects would have been different and required different semantic descriptions.

In sum, if meaning is functional role, then it is false that a representation would have had just the effects that it did have if its meaning had been different. Different meaning requires different functional role, and different functional role requires different causes and/or effects. So the Counterfactual Argument is unsound if meaning is functional role.

In terms of the original paradox, the point is that internal processors can be sensitive to both syntax and semantics. But how is this possible? Are the meanings of a gate's outputs dependent on the meanings of its inputs?

Of course not – and therein lies the fallacy of the original argument. Thinking of internal processors on the model of gates misleads. For a processor that is a *genuine intentional system* the difference between a representation's meaning one as opposed to truth or green, would involve differences in the internal part of the functional role of the representation. Not so for a gate. If we ask whether a representation would have had just the same behavioral effects had it had a different meaning, then for many differences in meaning, the answer will be yes for a genuine intentional system, though not for a simple primitive processor such as a gate. Indeed, the criterion of identity natural for a complex processor allows one to consider whether a given processor would have processed a representation differently had it had a different meaning. But the natural criterion of identity for a gate rules out the possibility of its processing differently (while remaining the same gate).

This point is easier to appreciate if we distinguish between autonomous and observer-relative meaning (Searle, 1980; Haugeland, 1980). Observer-relative meanings are inherited meanings, meanings

that intentional systems assign (e.g., to linguistic items or states of a machine). Autonomous meanings are meanings of representations or representational states of an intentional system – *for* that system. They get their meanings from their function in the system. The representations of gates have *only* observer-relative meanings. (My representations have autonomous meanings for me, but they can also have a variety of observer-relative meanings for others.) We can decide, if we like, that a '1' that is input to a gate means one, whereas an output '1' means Richard Nixon. We have a free hand. But autonomous meanings are not subject to whim in this way. It is for autonomous meanings that the point I have been making applies. Had an input symbol had a different autonomous meaning (for a genuine intentional system), then it could have had a different functional role and thus different effects. The trick of the argument that originally got us into trouble is to concentrate on the example of a gate in which only observer-relative meaning is relevant, making us forget that we do not have a free hand in this way with autonomous meaning.

Of course the autonomous meanings of my representations could have been different in *certain* ways without any change in the movements of my body. This is what is imagined in the famous "twin earth" examples – referential changes without changes in internal functional roles. But of course twin earth cases are elaborately artificial, and one should not conclude from them that had a representation meant something else, the same behavioral effects would have occurred nonetheless. The misleading effect of the gate example is to trick us into treating *all* hypothetical differences in meanings of representations as if they are twin earth cases. If we compare *and* gates whose '1's mean one versus truth, we are considering an analog of a twin earth case for gates. But for gates, twin earth cases come cheap. To read this cheapness of twin earth cases back onto genuine intentional systems is in effect to suppose that any old counterfactual situation in which references would be different is a twin earth case, and this is a bad mistake.

When I presented the Paradox of the Causal Efficacy of Content, I mentioned three premises: that intentional content is causally relevant to behavior, that intentional content reduces to meanings of internal representations, and that internal processors are sensitive to syntax, not meaning. Thinking of these as the premises, the argument is unsound because the third premise is false.

But when I spelled the argument out, I said that primitive processors are sensitive to syntax, not meaning. I then argued that the meaning of a representation cannot influence the behavior of a system of processors without influencing any of the particular processors themselves. So

if meaning ca
behavior of ;
cognitive sci
can be an int
the purposes
argument is s

5

Dretske (1988) is causally re
point has littl
been discussi
favor of the i
what they "i
production of
terms of the
Block, 1986, i
relevant to th
point – infor
sentations' fu
any actual be

Consider w
call it the fro
role of this re
its production
controlling th
functional rol
flies and their
causally releva
ancestors of t
the world –
production of
ancestor of th
content with r
influence of t
fitness on the
sort of influen
recruited information a
improving the
were replaced

if meaning can't influence the behavior of a gate, it can't influence the behavior of a system of gates. This reasoning is mistaken (given the cognitive science assumption that a computer – a system of gates – can be an intentional system – an assumption that I am accepting for the purposes of this discussion) and so understood this way, the original argument is sound but invalid.

5. Internal functional role and external content

Dretske (1988) has argued that the semantic content of a representation is causally relevant to behavior. This section will briefly note that his point has little to do with the problem of epiphenomenalism as I have been discussing it. I think of Dretske's considerations as counting in favor of the idea that the informational values of our representations – what they "indicate" about the world – are causally relevant to the production of the purely "internal" aspect of their functional roles. (In terms of the "two factor" version of functional role semantics [see Block, 1986, and further references there], the external factor is causally relevant to the production of the internal factor.) But – and this is my point – informational value can be causally responsible for our representations' functional roles without being involved in the "triggering" of any actual behavior (in the usual sense of 'behavior').

Consider whatever it is that the frog uses to represent flies – let's call it the frog's fly representer.⁷ There is an aspect of the functional role of this representer that is completely internal – mainly a matter of its production by flashes of movement on the frog's retina and its role in controlling the aim of tongue launchings. In addition to its internal functional role, this representer has informational content regarding flies and their locations. The point is that the latter plausibly has been causally relevant to the former. More exactly, the informational value of ancestors of this representer – what they have been indicating about the world – have been involved, or so one might suppose, in the production of the representer's current functional role. Perhaps an ancestor of the frog had an internal state that had some informational content with respect to ancestors of flies or other food on the wing, and influence of this state on tongue launching conferred extra inclusive fitness on the frog ancestors whose fly-information state had the right sort of influence on tongue launchings. We may speculate that evolution recruited a primitive motion detector that provided a modicum of information about winged bugs to guide the prefrog's tongue, thereby improving the prefrog's chances of a meal. As the bugs evolved into (or were replaced by) flies, the detector was turned by evolution to flies.

Then the fact that the frog's fly representer has been carrying information about flies has causally contributed to giving this representation the functional role that it has – being produced by retinal movement flashes and guiding zapping. More generally, the line of thought is that what our representations have been indicating about the world has had an influence – via evolution – on their having the functional roles that they have in our heads.

I do not wish to go into this reasoning in any detail, consider objections to it, or talk about the extrapolation from frogs to people. I want to point out only that even if the reasoning is entirely correct in its own terms, it does not show that the informational content of a representation is part of what is causally relevant to (in the sense of a "triggering" cause) the behavioral output that the representation causes, and so Dretske's proposal is not a solution to the problem being considered in this paper.

Suppose that the frog has a fly-word ('FLY'), the informational content of which has been causally relevant to the establishment of the internal functional role of 'FLY', including the guiding of the frog's tongue zapping behavior. Is this informational content thereby causally relevant to the production of, i.e., involved in the causation of, any particular zapping? This is the epiphenomenalism issue of this paper (applied to this case). The answer is obviously not. X can have causally promoted the pattern of $Y \rightarrow Z$ without in any way triggering the current token of Z. For X can have promoted $Y \rightarrow Z$ without now causing Y or enabling Y to cause Z. The informational content of 'FLY' does not causally contribute to the appearance of this token of 'FLY' in the frog's head. That is done by the fly that caused it. And once 'FLY' has appeared in the frog, the informational content does not enable or aid 'FLY' in producing a zapping. To dramatize the point, suppose that the current 'FLY' token is a *misrepresentation* caused, say, by a B-B. This fly token nonetheless indicates flies (a misrepresentation of a fly is still a representation of a fly). The history of correlation of 'FLY' tokens with flies has contributed to the functional role of 'FLY' tokens in the frog, but once that role is set, the past correlation is irrelevant to the process by which the B-B now produces the current 'FLY' token, which in turn produces the zapping that pops the B-B into the frog's gut. We can tell the whole mechanistic story about this causal process without saying anything about how the mechanisms that subserve it arose. And it is this former question that this paper is about: is content part of the causal process by which behavior is produced?

The plot so
ism and slay
the victory c
that functio
can win a b
the end be
crucial case
section is to
not have the
tentative.

In brief, i
consist in th
properties)
inputs and
non-function
not the func

To get at
that of a fut
which I mea
or other (say
one anothe
properties n
erties.) Con
ignoring the
provokes th
anger. The
tive, of hav
some proper
does the *pr
tiveness ca
bull is too
provoke the*

*Another c
property, th
first-order cl
x is dormitiv
(when x is ir
your noticing
falling asleep)*

6. Functional properties

The plot so far is: functionalism meets arguments for epiphenomenalism and slays them. It would be nice to stop here, but alas, the story of the victory of functionalism over epiphenomenalism is fiction. I argued that functionalism does defeat the Counterfactual Argument, but you can win a battle and still lose the war. Functionalism loses the war in the end because functional properties are causally inert in certain crucial cases. Or rather, I fear that all this is true. The point of this section is to raise a skeptical doubt. The issues are complex, and I do not have the space to explore them adequately. So my claims must be tentative.

In brief, my point is this. Functional properties are properties that consist in the having of some properties or other (say non-functional properties) that have certain causal relations to one another and to inputs and outputs. In the production of those outputs, it is the non-functional properties that are standardly the causally relevant ones, not the functional properties.

To get at the point, let's consider a slightly more general notion than that of a functional property, the notion of a second-order property, by which I mean a property that consists in the having of some properties or other (say first-order properties) that have certain causal relations to one another. (The greater generality here is just that second-order properties needn't involve inputs and outputs, as with functional properties.) Consider the bullfighter's cape. The myth (which we will accept, ignoring the inconvenient color-blindness of bulls) is that its red color provokes the bull; that is, redness is causally relevant to the bull's anger. The cape also has the second-order property of being provocative, of having some property or other that provokes the bull, of having some property or other that is causally relevant to the bull's anger. But does the *provocativeness* of the cape provoke the bull? Is the provocativeness causally relevant to the bull's anger? It would seem not. The bull is too stupid for that. The provocativeness of the cape might provoke the ASPCA, but not the bull.

Another example: consider dormititvity construed as a second-order property, the possession of some property or other (for example, a first-order chemical property) that is causally relevant to sleep. That is, x is dormitive = x has some property that is causally relevant to sleep (when x is ingested). If a dormitive pill is slipped into your food without your noticing, the property of the pill that is causally relevant to your falling asleep is a (presumably first-order) chemical property, not, it

would seem, the dormitivity itself. Different dormitive potions will act via different chemical properties, one in the case of Valium, another in the case of Seconal. But unless you know about the dormitivity of the pill, how could the dormitivity itself be causally relevant to your falling asleep?

Of course if you *do* know about the dormitivity, then it can be causally relevant to your sleep, just as the provocativeness of the cape can affect the ASPCA. In fact, there is a well-known phenomenon in which dormitivity *does* cause sleep, namely, the placebo effect. If a dormitive pill is so labeled, thereby causing knowledge of its dormitivity, this knowledge can cause sleep (though the truth and justification of the knowledge are of course causally irrelevant). So dormitivity can be causally relevant to sleep. (Indeed, there can be a dormitive pill that works without any first-order effect, a pill whose dormitivity requires its own recognition. Suppose I market a sugar pill as a dormitive pill, and it becomes popular and works well. I make a fortune and close my plant. Years later, when all my pills have been used up, one of my customers who had had years of sound sleep as a result of taking my pills finds out that they worked via the placebo effect and sues me. Surely I can point out that he was not cheated – the pills were genuinely dormitive, there was no false advertising.)

My claim is that second-order properties are not *always* causally relevant to the effects in terms of which they are defined.⁸ The only cases that I can think of in which second-order properties seem to be causally efficacious are those where an intelligent being recognizes them. That is why I keep mentioning "standard" cases, cases where a second-order property is defined in terms of an effect and that effect is produced without any recognition of the second-order property by an intelligent being. I add to the claim of the first sentence of this paragraph the more tentative claim that in these standard cases, the second-order property is causally inefficacious.

But how can it be that second-order properties are inert in some cases, efficacious in others? Think (temporarily) in terms of a nomist theory of causal relevance. If dormitivity of a pill were nomologically sufficient for the ingester getting cancer (Jerry Fodor keeps trying to convince me that such a thing could happen without recognition of dormitivity by an intelligent being), then (let us suppose) dormitivity would be causally relevant to cancer. But such causal relevance to cancer would not show or even suggest that dormitivity is *nomologically* sufficient for *sleep*. Hence we might have causal relevance to cancer but not to sleep: non-standard causal relevance without standard causal relevance. (More on nomological sufficiency in a moment.)

Second-
though it i
first-order,
procedure
a theory in
some sort
(In the cas
as theoreti
oretical te
leaving ou
define 'T'

So far, t
order prop
itive conse
are, a natu
taking suc
the causal
special sci
be that we
sleep, and
principles
two.

First, le
that conce
the effect
nomologic
sleep. The
being a w
and expa
dormitive
this defin
– and I t
sufficient
relation.
nomologic

Now, I
properties
for prope
my aunt's
tracks ch:
his is alw
nomologi

otions will act
im, another in
mitivity of the
to your falling

hen it can be
ss of the cape
henomenon in
bo effect. If a
its dormitivity,
ification of the
itivity can be
utive pill that
ity requires its
mitive pill, and
and close my
up, one of my
it of taking my
and sues me.
the pills were

always causally
ied.⁸ The only
ies seem to be
ing recognizes
cases where a
id that effect is
property by an
ntence of this
ard cases, the

inert in some
ms of a nomist
e nomologically
keeps trying to
recognition of
ose) dormitivity
al relevance to
is *nomologically*
ce to cancer but
standard causal
nt.)

Second-order properties involve having some properties or other, and though it is often helpful to think of the properties quantified over as first-order, actually they can be any properties at all. There is a general procedure (see Lewis, 1970) for defining a second-order property, given a theory in which the property plays a role so long as the theory allows some sort of a distinction between theoretical and observational terms. (In the case of a psychological theory, the distinction would be cashed as theoretical = mental, and observational = input/output.) If the theoretical terms are ' T_1 ' ... ' T_n ', we can write the theory as $T(T_1 \dots T_n)$, leaving out all mention of the observational entities. Then we can define ' T_i ' as follows: x has $T_i = EF_1 \dots EF_n [T(F_1 \dots F_n) \& x \text{ has } F_i]$.

So far, the case I've made for the limited causal inertness of second-order properties is based entirely on examples. But if the counterintuitive consequences of this causal inertness are as bad as I will claim they are, a natural response will be simply to live with rejecting my way of taking such examples. That is, if I am convincing later when I say why the causal inertness of the second-order commits us to a view of the special sciences that is hard to swallow, the reasonable response would be that we should just suppose that dormitivity *is* causally relevant to sleep, and provocativeness does affect the bull. So I will try to get at the principles that underlie our reaction to the examples. I can think of two.

First, let us return to the nomist conception of causal relevance. On that conception, second-order properties are not causally relevant to the effects in terms of which they are defined because they are not nomologically sufficient for those effects. Consider dormitivity and sleep. The relation between the two is more like the relation between being a widow and having had a husband than that between, say, heat and expansion. If a pill is dormitive in the following sense: x is dormitive iff x has some property that causally guarantees (this is where this definition differs from the one offered earlier) sleep if x is ingested – and I take the pill, it follows that I sleep. The fact that dormitivity is sufficient for sleep is perfectly intelligible in terms of this logical relation. What reason is there to suppose that there must *also* be a nomological relation between dormitivity and sleep?

Now, I am very much not saying that a logical relation between properties *precludes* a nomological relation. This is as much a fallacy for properties as for Davidsonian token events. Suppose dormitivity is my aunt's favorite property, and sleep is my uncle's, and that my uncle tracks changes in my aunt's favorite property, changing his own so that his is always entailed by hers. Then dormitivity and sleep will be *both* nomologically and logically related. Logically related under one set of

descriptions, nomologically under another. My point is not that a logical relation precludes a nomological relation, but rather that the logical relation between dormitivity and sleep tells us perfectly well why dormitivity involves sleep. There would have to be some *special* reason to postulate a nomological relation as well, and since the story about my aunt and uncle wasn't true, I don't see any such special reason. The point that this example is meant to make is that it would be amazing if there was *always* some special reason why a second-order property was nomologically related to the effects in terms of which it is defined.

This consideration is based on the fact that second-order properties are defined in terms of effects. A second argument is based on a different feature of second-order properties, the quantification involved in them.

Supposing that provocativeness provokes the bull would be supposing a strange sort of overdetermination of the bull's anger. Of course overdetermination does sometimes happen. (The placebo effect is an example.) But to suppose that it always happens would be to suppose a bizarre systematic overdetermination. Whenever we have a first-order causal relation we can always define a second-order property on the model of provocativeness, and so every first-order causally relevant property would jointly determine its effect together with a second-order property. Indeed, the procedure iterates (for there is a third-order property that consists in the possession of a second-order property that is causally relevant to the effect), and so whenever there is one causally relevant property there would be an infinity of them. Even if the first-order property is causally sufficient for the effect, there would still be an unending series of other causally relevant properties. And we can define causally sufficient higher-order properties that would also be causally sufficient by the same reasoning.

The relation between second-order properties and the effects mentioned in their definitions is a "fork" relation, a *bit* like the one discussed earlier in the selection on correlation and causation. Both heat and electricity are conducted by free electrons in metals. Rising velocity of free electrons

- I. Is responsible for rising thermal conductivity, and thus causally relevant to the explosion
- II. Is responsible for the epiphenomenal rising electrical conductivity.

Similarly, there is a first-order chemical property of Seconal that

- I. Is causally relevant to sleep
- II. "Generates" an epiphenomenal second-order property of possessing some property that is causally relevant to sleep.

In other w
explosion v
conductivity
ing an epip

The pict
analogy ju
relations n
different, a
(limited) ep
used to pr
chemical pi
tivity illustr
nalism' and
gument also
overdeterm
coincidence
is always a
coincidence
causally effi
causally effi
the problem

In the ei
relevance i
favor count
as earlier in
nalism, go j
cal theory.
though its
approach.

Let me s
property is
occurs; my
relevant to
second-orde
defined (anc
of them tak
second-orde
would not c
epiphenome
tion argume
order prope
effects. The
of content

hat a logical
t the logical
ly well why
special reason
ory about my
reason. The
e amazing if
property was
defined.
r properties
based on a
ion involved

be supposing
. Of course
effect is an
to suppose a
a first-order
perty on the
ally relevant
econd-order
third-order
perty that
one causally
Even if the
e would still
And we can
uld also be

effects men-
ke the one
sation. Both
etals. Rising
nus causally
I conductiv-
nal that
y of possess-

In other words, just as the rising velocity of free electrons causes the explosion while engendering an epiphenomenal increase in electrical conductivity, so the chemical property causes the sleep while engendering an epiphenomenal second-order property of dormitivitiy.

The picture just sketched is attractive, but hardly compelling. The analogy just mentioned is far from perfect. The two engendering relations mentioned in the preceding paragraph are certainly very different, and besides the analogy only holds if I am right about the (limited) epiphenomenality of second-order properties, so it cannot be used to prove that epiphenomenality. (Incidentally, the fact that the chemical property "engenders" dormitivitiy rather than causing dormitivitiy illustrates the difference between what I am calling 'epiphenomenalism' and traditional epiphenomenalism.) The overdetermination argument also is far from convincing. We are normally reluctant to accept overdetermination because it is wrong, other things equal, to postulate coincidences. If a man dies by drowning, we cannot suppose that there is always another cause of death as well, say, shooting. But no such coincidence would be involved in the series of higher-and-higher-order causally efficacious properties I mentioned. If accepting such a series of causally efficacious properties is a price that must be paid for avoiding the problems to be mentioned, it can be paid.

In the end, the argument based on nomological theories of causal relevance is the best one. However, we can hardly expect those who favor counterfactual theories of causal relevance to be convinced. *Here, as earlier in the paper, the lesson is that if you want to avoid epiphenomenalism, go for a counterfactual theory of causal relevance, not a nomological theory.* This is, I suppose, the main positive point of the paper, though its significance depends on the fate of the counterfactual approach.

Let me sum up the skeptical thesis. Suppose that a second-order property is instantiated, and the effect in terms of which it is defined occurs; my claim is that the second-order property needn't be causally relevant to the effect. I have mentioned one type of case in which a second-order property can affect the effects in terms of which it is defined (and other things as well), namely, when intelligent recognition of them takes place, as in the placebo effect. Accepting other sorts of second-order effects (e.g., causal relevance of dormitivitiy to cancer) would not change my claim, since the arguments I gave for the limited epiphenomenality – the nomological argument and the overdetermination argument – were restricted to the causal relevance of a second-order property on the effects in terms of which it is defined, not other effects. The epiphenomenalism I am worried about is not total inertness of content properties, but whether, for example, the content of my

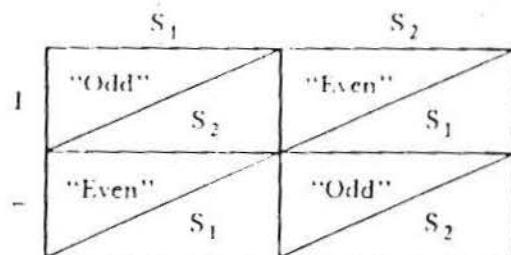


Figure 5

desire for an ice cream cone is causally relevant to my going to the ice cream shop.⁹

At this point, I shall change gears abruptly. Although, as I've said, the case I've made for the (limited) ineffectiveness of the second-order property is far from conclusive, I think I have said enough for it to be taken seriously. What I now propose to do is simply *assume* that second-order properties are not always causally relevant to the effects in terms of which they are defined, and go on to discuss the consequences of this idea.

The skeptical point of this section depends on the fact that the mechanisms that manipulate genes and the mechanisms that manipulate the internal representations in the brain are not intelligent, and so cannot recognize second-order properties. Hence these are just the kinds of cases in which it would seem that the second-order properties are ineffectual with respect to the effects in terms of which they are defined. Likewise for other special sciences whose properties are plausibly second-order.

By way of an example, consider a simple finite automaton, the one specified by the machine table of Figure 5. What the table requires of any machine that it describes is that it have two states, S_1 and S_2 , such that when the machine is in S_1 and sees a '1' it cries 'odd' and goes into S_2 ; and when the machine is in S_2 and sees a '1', it cries 'even' and goes back into S_1 . And when the machine is in S_1/S_2 and sees no input ('-' symbolizes the null input), it cries 'even'/ 'odd' (respectively), staying in the same state. Thus the machine tells us whether it has seen an odd or even number of '1's (though not entirely successfully, since it is under the impression that 0 is an even number).

Now suppose we have a machine on the desk in front of us that satisfies Figure 5 (i.e., is described by the table) and is in S_1 ; it sees a '1', and so it cries 'odd'. Let us ask: is the property of being in state S_1 (of a machine of the type of Figure 5) causally relevant to the production of the cry of 'odd'? If the points made about dormitivity and provocativeness are right, it would seem not. If we think of S_1 as a property, it is

the property related to on 5. These other a gears-and-computer im tronic imple causally relev such implem

The upshe what compu implementat strange conc comparably tional scienc

Consider plausibly fun tance and actual chara dominant ov of the disti there are n property of chemical pr via the proc tional prop processors ties such a processors properties which they gene.¹⁰

In sum, inconsisten

1. Speci scienc
2. Speci
3. Funct effect

I have ju us is a d epiphenon

the property that consists in having two other properties that are related to one another and to inputs and outputs as specified in Figure 5. These other properties would be mechanical properties in the case of a gears-and-pulleys implementation, electronic properties in a usual computer implementation, and so on. If we are dealing with an electronic implementation, then it is the electronic properties that are causally relevant to the cry of 'odd', not the property of possessing some such implementation property.

The upshot is that *computational properties have no causal relevance to what computers do*, these effects being entirely the product of the implementations of the computational properties. This is indeed a strange conclusion, and if I am right about second-order properties, comparably strange conclusions could be reached about other functional sciences.

Consider the property of being a brown-hair gene. This property is plausibly functionally characterizable in terms of its patterns of inheritance and interaction with other genetic properties in determining actual characteristics, for example, in the fact that brown-hair genes are dominant over blond-hair genes. Such second-order properties are part of the distinctive conceptual apparatus of Mendelian genetics. But there are no genetic mechanisms that detect or are sensitive to the property of being a brown-hair gene. Genetic mechanisms detect biochemical properties, and the property of being a brown-hair gene arises via the processing based on biochemical properties. Similarly, computational properties such as S_1 mentioned above arise via the activity of processors that detect syntactic properties, not computational properties such as S_1 . And semantic properties arise via the activity of processors that detect syntactic properties. All these second-order properties are epiphenomenal with respect to the effects in terms of which they are defined, such as brown hair in the case of the brown-hair gene.¹⁰

In sum, we want to maintain all of the following, but they are inconsistent:

1. Special science properties are causally relevant to the effects those sciences predict and explain.
2. Special science properties are often functional.
3. Functional properties are standardly causally irrelevant to the effects they are defined in terms of.

I have just argued for 3. If I am right, then we have what for many of us is a dilemma: we must either abandon functionalism or accept epiphenomenalism. But am I right about 3? Recall that the case for 3

depended mainly on the nomological approach to causation, and perhaps the way out is to reject that approach. Before accepting that conclusion, however, it would be best to examine another way out.

7. Because

The claim that second-order properties are standardly causally irrelevant to the effects in terms of which they are defined seems to fly in the face of a variety of commonsense and scientific facts, for example, true uses of the following:

- I slept because I took a sleeping pill.
- My muscle tension evaporated because I took a muscle relaxant.
- The vase broke because it was fragile.
- I have blue eyes because I have two blue-eye genes

I shall argue that such facts only seem incompatible with the (standard) causal inefficacy of the second order because of two ambiguities that easily escape notice.

Explanation and causation

One ambiguity is that between explanatory and causal uses of 'because'. It is easy to be unclear about whether a 'because' statement is true in an explanatory or causal sense of 'because'. If the dispositional/functional terms in the sentences listed above are understood to refer to second-order properties, then the sentences are only true on an explanatory, not a causal, reading of 'because' – if the line of thought of the last section is correct.

Audiences who have heard this chapter read as a paper have often objected that although there is a clear distinction between explanation and causation, the distinction between the *causal-explanatory relevance* and *causal relevance of properties* that I rely on here is a distinction without a difference. Here is the issue: The kind of explanation involved here is causal explanation, which is one of a number of species of explanation. So the kind of explanatory relevance of properties at issue is causal-explanatory relevance. And what is the distinction between causal-explanatory relevance of properties and causal relevance of properties?

Here is a way of seeing that there is a difference and what it is: notice that we can causally explain a case of sleep via appeal to the second-order property of dormitivity. I fell asleep because I took a pill that had the property of having some property or other that causes sleep. This is

causal-explanatory of my falling second-order appeal to do Dormitivity causally relevant appeal to a function of sleep chemical property second-order in the possession cause it involves property can causally effective first-order or for the standard

Functional terms One construal dormitivity = one equally just. This is the natural dormitive virtue construal is a situation that picks former constituent context. 'I took the same second different first-order properties. For different from the lotteries (Lewis)

To sum up relevant to situations the 'because' is ascribed to the

David Lewis Smart's idea Functional terms

i, and per-
epting that
ay out.

sally irrele-
o fly in the
imple, true

nt.

the (stan-
mbiguiities

'because'.
t is true in
onal/func-
to refer to
on an ex-
thought of

have often
xplanation
y relevance
distinction
ination in-
of species
roperties at
inction be-
relevance

it is: notice
ne second-
ill that had
ep. This is

causal-explanatory because it rules out alternative causal explanations of my falling asleep – for example, that I was grading papers. But second-order properties being (standardly) causally inefficacious, the appeal to dormitivity is not an appeal to a causally relevant property. Dormitivity is causal-explanatorily relevant to sleep without being causally relevant to sleep. Why? The appeal to dormitivity involves an appeal to a property that is genuinely causally relevant to the production of sleep, namely, the unnamed property (presumably a first-order chemical property) whose existence is mentioned in the analysis of the second-order property, the one that the dormitivity of the pill consists in the possession of. Dormitivity has causal-explanatory relevance because it involves a causally efficacious property. A causally inefficacious property can nonetheless be causal-explanatory if it "brings in" a causally efficacious property. But the causally relevant property is the first-order one, not the second-order one (assuming that my argument for the standard causal inertness of the second-order was right).

Ambiguity of functional terms

Functional terms are ambiguous. Let's take 'dormitivity' as an example. One construal is the second-order one much discussed here already – dormitivity = *possessing a property that causes sleep*. The other construal, one equally justified in ordinary usage, is *the property that causes sleep*. This is the natural reading of 'dormitive virtue' in the assertion that the dormitive virtue of Seconal is causally responsible for sleep. This latter construal is one in which 'dormitivity' is analyzed as a definite description that picks out the first-order property that is quantified over in the former construal. The two construals are very different in their relation to context. 'Dormitivity', on the former construal, always picks out the same second-order property. But on the latter construal, it picks out different first-order properties in the case of different types of sleeping potions. For the property that causes sleep in the case of Seconal is different from the property that causes sleep in the case of Valium. Similarly, 'the winning number' picks out different numbers in different lotteries (Lewis's example).

To sum up: In the second-order sense, dormitivity is not causally relevant to sleep; in the first-order sense, it is. And the plausibility of the 'because' claims mentioned at the outset of this section can be ascribed to the first-order interpretation.¹¹

David Lewis has long advocated a form of functionalism based on Smart's idea of topic neutral analyses of mental terms (Lewis, 1972). Functional terms are defined by Lewis in the latter of the two manners

mentioned above – in terms of definite descriptions. Many functionalists adopt a functional state identity thesis, one that identifies mental states with second-order properties. Lewis, by contrast, says that mental states are *functionally specified* brain states. Many functionalists take pain to be a second-order state, but Lewis takes pain to be the first-order brain state that plays the role characterized by the second-order state. Where some functionalists identify mental states with states that consist in causal roles, Lewis takes mental states to be the entities that have the causal roles. The issue is about how to regiment the language of the mental (and other theoretical languages).

Now some may embrace Lewis's first-order reading of functional terms of the special sciences as a way of having their cake and eating it too. On Lewis's construal, it is correct to say "Dormititory is causally relevant to sleep," "Content is causally relevant to behavior," and the like.¹² If we accept Lewis's reading, we can preserve the talk that would be appropriate if special science properties were causally efficacious. So it seems that we can avoid epiphenomenalism, and not have to give up functionalism. And there is no need to reject a nomological theory of causal relevance.

Unfortunately, the Lewis solution (as I shall call it) is only cosmetic. To construe the functional terms of a special science in Lewis's manner is to construe them as picking out "lower level" properties, properties of an *implementation* science. 'Believing that grass is green' on the Lewis construal picks out a *physiological* property. (In us, that is; in the case of a hypothetical intelligent computer, it would pick out an electronic property.) And in a certain sense, 'believing that grass is green' does not pick out a *psychological* property; that is, it does not pick out a property that is part of the distinctive conceptual apparatus of psychology. On Lewis's construal, 'brown-hair gene' (a term used by Mendelian geneticists long before the advent of molecular biology) picks out a piece of DNA, the concept of which would have been utterly alien to Mendel. Consider the computational state S_1 of Figure 5. I said it was a computational state, and indeed the computational definition was given in the text. But on Lewis's construal, ' S_1 ' picks out an electronic state or a mechanical state, or some other implementation state, depending on the way the automaton is constructed. ' S_1 ' would be defined by Lewis's method as *the (contextually relevant) state that has the role specified in the table of Figure 5*. So ' S_1 ' does not pick out a computational state at all.

Of course, if a "psychological property" is just a property picked out by a psychological term, then 'believing that grass is green' does pick

out a psychological property. Pickwickian the conceptual part of the traditional pro nothing about the status of electric

My point properties properties strual does not state S_1 b of "Even! tional stat the causa larly, since psycholog believing of the me

Constru make the distinctive way of n Lewis's n But 'havi ' S_1 ' picks tition state Computa sciences, to them ences. T sciences conceptu special properti effects in

In su epiphen allows u

Many functional-
identifies mental
, says that mental
unctionalists take
pain to be the
d by the second-
states with states
to be the entities
to regiment the
es).

ng of functional
ake and eating it
tivity is causally
behavior," and the
e talk that would
lly efficacious. So
t have to give up
logical theory of

is only cosmetic.
Lewis's manner
erties, properties
s green' on the
us, that is; in the
ld pick out an
ng that grass is
it is, it does not
epatial apparatus
(a term used by
olecular biology)
ave been utterly
f Figure 5. I said
tional definition
' picks out an
implementation
ed. 'S₁' would be
state that has the
not pick out a

property picked out
'green' does pick

out a psychological property, and likewise 'brown-hair gene' does pick out a Mendelian property and 'S₁' does pick out a computational property. But these construals of "psychological property", etc., are Pickwickian. Psychological properties, on this construal, are not part of the conceptual apparatus of psychology; Mendelian properties are not part of the conceptual apparatus of Mendelian genetics; and computational properties are properties that a computer scientist may know nothing about, properties that are really part of the conceptual apparatus of electronics or mechanics or hydraulics.

My point is this: You do not vindicate the causal efficacy of the properties of a special science by construing its terms as referring to the properties of other sciences, implementation sciences. The Lewis construal does preserve a way of talking that we want. We can speak of the state S₁ being causally relevant to the production of the machine's cry of "Even!" But the cost is that we no longer take S₁ to be a computational state (except in a Pickwickian sense). So a Lewis vindication of the causal efficacy of the computational is certainly Pickwickian. Similarly, since believing that grass is green is a physiological, not a psychological, state, the preservation of talk of the causal efficacy of believing that grass is green is no real vindication of the causal efficacy of the mental.

Construing terms of the special sciences in Lewis's manner does not make the second-order properties I have been talking about any less distinctive of their special sciences. Rather, it merely forces a different way of referring to them. 'Believing that grass grows', construed in Lewis's manner, picks out a physiological state, not a functional state. But 'having the belief that grass grows' picks out the functional state. 'S₁' picks out an electronic or mechanical or some other implementation state, but 'being in S₁' picks out a genuine computational state. Computational states are what are distinctive about the computational sciences, and Lewis's way of talking does not prevent us from referring to them or make them any less distinctive of the computational sciences. The same holds for the functional properties of all the special sciences for which functional properties are part of the distinctive conceptual apparatus. These properties remain distinctive of these special sciences even if we construe terms à la Lewis, and these properties remain causally inert (that is, causally inert with respect to effects in terms of which they are defined).

In sum, moving to the Lewis way of talking does not solve the epiphenomenalism problem, but only provides a way of talking that allows us to avoid facing it. The basic picture provided by the Lewis way

of talking is the same as the one I have been trying to avoid: special sciences whose distinctive properties are causally inert but that can be used to pick out causally efficacious properties of lower-level sciences.

This brings me to a second objection to the Lewis solution. Second-order properties in the sense discussed here are properties that consist in the having of properties with such and such relations to one another. As I mentioned earlier, the properties quantified over need not be first-order. In the case of our mental properties construed as second-order properties, the quantified properties are physiological. Physiological properties are biological properties, and biological properties generally are prime candidates for functional analysis. Thus 'believing that grass grows' construed in the Lewis manner, since it picks out a physiological property, *may pick out another functionally individuated property*. Perhaps biological terms themselves pick out first-order properties, but perhaps not. Perhaps only the most fundamental of physical properties are genuinely first-order. In that case, even much of physics would be epiphenomenal! Or perhaps not even the most fundamental physical properties are first-order. (Maybe there isn't any most fundamental level.) Perhaps science is functional "all the way down." (Recall that there is a case made by Lewis himself for functional analysis of all theoretical terms.) This is not a matter that is easily resolved, but without a resolution that rules out the "functional all the way down" option, we cannot rely on construing mental language in Lewis's manner to rescue us from epiphenomenalism.

Construing special science terms in Lewis's manner makes epiphenomenalism easier to stomach, but epiphenomenalism with controlled heartburn is still epiphenomenalism. Perhaps the solution is to reject functionalism, but if functionalism is rejected, we must begin anew to confront the arguments for epiphenomenalism in the computer model all over again. Functionalism at least has the virtue of disposing of epiphenomenalism on a counterfactual understanding of causal relevance. However, if we prefer the nomological approach to causal relevance, I am doubtful that there is any way of avoiding epiphenomenalism.

Notes

¹ There are, however, important differences that I don't have the space to go into. Briefly, it is more plausible that content supervenes on the neural level than that content supervenes on the "syntactic" level. (More about the meaning of 'syntactic' here shortly.) An appeal to supervenience of content on lower levels can be used to attempt to avoid the epiphenomenalism argument.

² The idea d
See also D
first appear
causal inte
among the
if it encou
symbols, ar
meanings a
identifying
neurologic:

³ This has lo
been challe
see Kim (1st
(1980a) arg
and Blackb

⁴ Alternative
temperatur
explode wi
part will co
to be unco
the explosio
conductivity
condition is

⁵ See Block (

See Fodor (

⁶ LePore and
Counterfac
syntactic or
have cause
meanings a
stances, the
by a differe
screen off t
meanings se
the syntacti
that is absu
Counterfact
symmetrical
other, so th
club against

This issue
The LePore
particular e
in my head
behavior. B
agree that S
meaning at
Bruce even
Certainly no
all, then it
Had the ori
we may su
Bruce-like b
ated causall

- oid: special
that can be
el sciences.
on. Second-
that consist
ne another.
eed not be
as second-
Physiologi-
ties gener-
lieving that
icks out a
'individuated'
order prop-
of physical
of physics
fundamental
lost funda-
n." (Recall
alysis of all
olved, but
way down"
wis's man-
- es epiphe-
controlled
s to reject
n anew to
iter model
sposing of
usual rele-
to causal
phenome-
- to go into.
el than that
of 'syntactic'
be used to
- 2 The idea described here was, as far as I know, first articulated in Fodor (1975, 1980). See also Dennett (1981) in which the terms 'semantic engine' and 'syntactic engine' first appear, and Newell (1980) and Pylyshyn (1984). I spoke of a correlation between causal interactions among symbolic structures in our brains and sensible relations among the meanings of the symbol structures. This way of speaking can be misleading if it encourages the picture of the neuroscientist opening the brain, just *seeing* the symbols, and then figuring out what they mean. Since syntactic objects (and perhaps meanings as well) are functional entities (though different types of functional entities), identifying either of them in the brain will require considerable appreciation of neurological functioning.
 - 3 This has long been a central dogma of the computational point of view, and has long been challenged by some philosophers. For arguments against the claim stated here, see Kim (1972), Richardson (1979), Erc (1983), and Patricia Churchland (1986). Block (1980a) argues against Kim, Patricia Kitcher (1980, 1982) argues against Richardson, and Blackburn (forthcoming) argues against Erc and Churchland.
 - 4 Alternatively, we could raise the thermal conductivity of the rod by raising its temperature, in which case part of the additional heat that causes the bomb to explode will come from the heat source that raises the temperature of the rod, and part will come from the fire via the increased thermal conductivity of the rod. I take it to be uncontroversial that rising thermal conductivity of the rod is causally relevant to the explosion. Further, this case fits the nomist model. For in this setup, rising thermal conductivity, *ceteris paribus*, is sufficient for an explosion, (and the *ceteris paribus* condition is satisfied).
 - 5 See Block (1986, 1987) for other reasons for taking functional role theories seriously. See Fodor (1987a), Chapter 3, and Schiffer (1987), Chapter 2, for opposing arguments.
 - 6 LePore and Loewer (1987) give an objection (in effect) to Sosa's (1984) version of the Counterfactual Argument. Putting the matter in my terms, they agree that if a syntactic object with a certain meaning produces an output, the syntactic object would have caused the same output even if it had had a different meaning. But they add that meanings are lawfully related to behavior in some circumstances, so in such circumstances, the meaning would have caused the same output even if it had been carried by a different syntactic object. If the first counterfactual shows that syntactic forms screen off meanings from causal relevance, then the second shows the converse, that meanings screen off syntactic forms from causal relevance. But if both are true, both the syntactic form and the meaning of a representation are epiphenomenal, and since that is absurd, the obvious conclusion is that something is wrong with the form of the Counterfactual Argument. According to LePore and Loewer, semantics and syntax are *symmetrically* related to the production of behavior; if one is epiphenomenal, so is the other, so the Counterfactual Argument cannot use the causal relevance of syntax as a club against the causal relevance of semantics.

This issue deserves much more detailed treatment, but I can only summarize briefly. The LePore and Loewer argument fails to take into account the distinction between a particular event token, and *some* token of that type. Suppose I have a syntactic object in my head, Sally, with certain syntactic and semantic properties that produces behavior, Bruce. As I mentioned in the last paragraph, LePore and Loewer and I agree that Sally would have caused Bruce even if it had had a different meaning or no meaning at all. But can we make the *symmetrical claim* that Sally would have caused Bruce even if it had had a different syntactic identity or no syntactic identity at all? Certainly not! If Sally had had a different syntactic identity or no syntactic identity at all, then it simply would not be the syntactic object I originally picked out as 'Sally'. Had the original meaning been "carried" by a syntactic object not identical to Sally, we may suppose that the result would have been a Bruce-like behavior. But this Bruce-like behavior would not be identical to Bruce – given that events are individuated causally, as all good Davidsonians (such as LePore and Loewer) believe. Bruce's

identity requires being caused by Sally. LePore and Loewer may feel that the way out is to talk not about the production of Bruce, but rather of Bruce-like events – types, not tokens. (There is some hint of this in their paper.) But the screening argument only works with tokens. What makes us want to say that the syntactic screens off the semantic is that the very same bit of behavior would have been produced even if the syntactic event that produced it had had a different meaning. (See Enc [1986] for a dissection of another fallacy that falls afoul of the type/token distinction together with causal individuation.) LePore and Loewer are right that there is a *nomological symmetry* between syntax and semantics. That is, there are circumstances in which both the semantic and syntactic properties of an event are nomologically sufficient for a certain type of behavior. But what is at issue here is counterfactual symmetry rather than a nomological symmetry, and this they have not argued for.

- 7 Note that it is no part of the "Language of Thought" view to insist that the frog has a language of thought. The argument for a language of thought in humans depends on the productivity and systematicity of human thought, and this argument will not apply to creatures whose thought is not at least as productive and systematic. See the epilogue of Fodor (1987a).
- 8 My point differs from the usual skepticism about the causal efficacy of dispositional properties in this respect, and also in that it is about a category of properties – second-order properties – of which dispositional properties are only a special case.
- 9 After writing this chapter, I read Jackson and Pettit (1988), which claimed that second-order properties are inefficacious before I did. However, Jackson and Pettit say that second-order properties are inert, not just with respect to the effects in terms of which they are defined, but all effects. They don't give any of the arguments used here, though Jackson tells me (in correspondence) that the overdetermination argument is essentially consonant with their way of looking at things. They go further, committing themselves, I think, to the claim that all "multiply realizable" properties are causally inert. Thus if an increase in temperature of a gas inside a sealed glass container results in the glass shattering, they take the fact that there are many configurations of molecules that could realize this process to show that the increase in temperature did not cause the shattering, but rather the cause was the specific molecular interactions. They do not note that the same point applies to the molecular interactions, which are themselves multiply realizable in terms of interactions among electrons, protons, neutrons, etc. And of course, some of these particles are perhaps realizable in different ways by still smaller particles. The upshot of their position would be that the only genuine causality inheres at the level of basic physics. But what if there is no such level? This is a real physical possibility. (See Dehmelt, 1989, 1990.) Their position dictates that we do not now know for sure whether there are any causally efficacious properties at all. Jackson tells me that they are prepared to locate all genuine causality at the basic physics level, so long as one distinguishes between two types of causal relevance: the notion used here and the notion of causal relevance that corresponds to their idea of a causal programming explanation. The idea of causal programming can be illustrated by their example of an explanation of why trees grow faster in Melbourne than in Canberra: Why? Because there are more frosts in Canberra. There being more frosts in Canberra causally programs there being faster growth in Melbourne. They hold that second-order properties (and other multiply realizable properties) can be causally relevant to something in this causal programming sense, but not in the sense appropriate only to basic physics.
- 10 For the purposes of this example, I am ignoring the issue of whether there is anything at the molecular level that corresponds very well to the gene. See Philip Kitcher (1982).
- 11 See Block (1980a) for a more detailed discussion of the difference between the first- and second-order interpretations.
- 12 Block (1986), pp. 668–9, takes this way out, as does Jackson and Pettit (1988).

- Block, Ned, 1
Mass.: Ha
1980b, "Wh
1986, "Adve
eds., Mid
Minnesota:
1987, "Func
Society, su
Churchland, P
Dehmelt, Han
sion?" *Pr*
1990, "Expe
Science, 2,
Dennett, Dani
ed., *Reduc*
Dretske, Fred,
bridge, M.
Enc, Berent, 1
80, 279–98
1986, "Esse
venience,
Studies in
403–26.
Fodor, Jerry,
University
1980, "Meth
nitive Psyc
1987a, *Psych*
1987b, "Mak
Loewer (1)
In press, "M
Haugeland, Jol
ioral and I
1981, *Mind* 1
Horwich, Paul,
Jackson, Frank
Mind, 97,
Kim, Jaegwon,
Identity Tl
1984, "Epiph
eds., Midw
Minnesota

Bibliography

- Block, Ned, 1980a. *Readings in Philosophy of Psychology*, Vol. I. Cambridge, Mass.: Harvard University Press.
- 1980b. "What Is Functionalism?" In Block (1980).
1986. "Advertisement for a Semantics for Psychology." In P. A. French et al., eds., *Midwest Studies in Philosophy*, Vol. X. Minneapolis: University of Minnesota Press, 615-78.
1987. "Functional Role and Truth Conditions," *Proceedings of the Aristotelian Society*, supplementary volume, 61, 157-81.
- Churchland, P. S., 1986. *Neurophilosophy*. Cambridge, Mass.: MIT Press.
- Dehmelt, Hans, 1989. "Triton, . . . electron, . . . cosmon, . . . : An infinite regression?" *Proceedings of the National Academy of Sciences*, 86, 8618.
1990. "Experiments on the structure of an individual elementary particle," *Science*, 247, 539-45.
- Dennett, Daniel, 1981. "Three Kinds of Intentional Psychology." In R. Healy, ed., *Reduction, Time and Reality*. Cambridge: Cambridge University Press.
- Dretske, Fred, 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Mass.: MIT Press.
- Enc, Berent, 1983. "In Defence of the Identity Theory," *Journal of Philosophy*, 80, 279-98.
1986. "Essentialism with Individual Essences: Causation, Kinds, Supervenience, and Restricted Identities." In P. A. French et al., eds., *Midwest Studies in Philosophy*, Vol. XI. Minneapolis: University of Minnesota Press, 403-26.
- Fodor, Jerry, 1975. *The Language of Thought*. Cambridge, Mass.: Harvard University Press.
1980. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," *Behavioral and Brain Sciences*, 3, 417-24.
- 1987a. *Psychosemantics*. Cambridge, Mass.: MIT Press.
- 1987b. "Making Mind Matter More." This is Fodor's reply to Lepore and Loewer (1987). The abstract appeared in *Journal of Philosophy*, 84, 642.
- In press. "Making Mind Matter More," *Philosophical Topics*.
- Haugeland, John, 1980. "Programs, Causal Powers, and Intentionality," *Behavioral and Brain Sciences*, 3, 432-3.
1981. *Mind Design*. Cambridge, Mass.: MIT Press.
- Horwich, Paul, 1987. *Asymmetries in Time*. Cambridge, Mass.: MIT Press.
- Jackson, Frank, and Philip Pettit, 1988. "Functionalism and Broad Content," *Mind*, 97, 387, 381-400.
- Kim, Jaegwon, 1972. "Phenomenal Properties, Psychophysical Laws, and the Identity Theory," *Momist*, 56, 177-92. Partially reprinted in Block (1980).
1984. "Epiphenomenal and Supervenient Causation." In P. A. French et al., eds., *Midwest Studies in Philosophy*, Vol. IX. Minneapolis: University of Minnesota Press; 257-70.

- Kitcher, Patricia, 1980. "How to Reduce a Functional Psychology," *Philosophy of Science*, 47, 134-40.
1982. "Genes, Reduction and Functional Psychology," *Philosophy of Science*, 49, 633-6.
- Kitcher, Philip, 1982. "Genes," *British Journal for the Philosophy of Science*, 33, 337-59.
- LePore, E., and B. Loewer, 1987. "Mind Matters," *Journal of Philosophy*, 84, 630-41.
- Lewis, David, 1970. "How to Define Theoretical Terms," *Journal of Philosophy*, 67, 427-46.
1972. "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy*, 50, 249-58. Reprinted in Block (1980).
- Minsky, Marvin, 1967. *Computation*. Englewood Cliffs, N.J.: Prentice-Hall.
- Newell, Alan, 1980. "Physical Symbol Systems," *Cognitive Science*, 4, 135-83.
- Putnam, Hilary, 1960. "Minds and Machines." In S. Hook, ed., *Dimensions of Mind*. New York: New York University Press.
1967. "Psychological Predicates." In W. H. Capitan and D. D. Merrill, eds., *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press. Retitled "The Nature of Mental States" in Putnam's collected papers. Reprinted in Block (1980).
- 1973, 1975. "Philosophy and Our Mental Life." *Mind, Language and Reality: Philosophical Papers*, Vol. 2. Cambridge: Cambridge University Press. Reprinted in Block (1980), and in somewhat different form, in Haugeland (1981). Originally published in *Cognition*, 2 (1973), with a section on IQ that has been omitted from all of the reprinted versions.
1988. *Representation and Reality*. Cambridge, Mass.: MIT Press.
- Pylshyn, Zenon, 1984. *Computation and Cognition: Issues in the Foundations of Cognitive Science*. Cambridge, Mass.: MIT Press.
- Richardson, Robert, 1979. "Functionalism and Reductionism," *Philosophy of Science*, 46, 533/558.
- Schiffer, Stephen, 1987. *Remnants of Meaning*. Cambridge, Mass.: MIT Press.
- Searle, John, 1980. "Searle's Reply to Critics of 'Minds, Brains and Programs,'" *Behavioral and Brain Sciences*, 3, 450-7.
- Sosa, Ernest, 1984. "Mind-Body Interaction and Supervenient Causation." In P. A. French et al., eds., *Midwest Studies in Philosophy*. Vol. IX. Minneapolis: University of Minnesota Press, 271-82.

F
Scientific realistic produ causal struc adoption o them, and causal struc the case fo apparently involve refre tures; the c to be theory 1988a, 198 Putnam 197 that appeal

A variety of theory-dep istic and p various sci edge but concepts a linguistic c theory-inde