Jason Bridges
University of Chicago

To commit Euthyphro's fallacy is to endorse a pair of incompatible explanations, one

constitutive and the other causal. Asked to explain the nature of piety, Euthyphro hazards that

being pious consists in being an object of the gods' love. But asked what causes the gods to love

what they do, he holds with the commonsensical thought that the gods love pious people because

they are pious. As Socrates points out (and for reasons we shall shortly rehearse), Euthyphro

cannot have it both ways. To hold that one's god-belovedness is constitutive of one's status as a

pious person is to rule out its being one's piety that prompts the gods' affection. More generally,

we commit the fallacy when we hold of two properties f and g both of the following: possession

of f constitutes possession of g, and possession of g causes possession of f.[1]

In the course of pursuing the familiar philosophical project of reducing one type of fact to

another, philosophers are often led to analyses that run opposite the grain of our ordinary view of

the phenomena in question. The most striking instances of this are cases of what Kripke calls

"inversion of a conditional", in the central species of which what we are pre-philosophically

inclined to regard as a cause/effect relation is "inverted" such that the presence of the 'effect' is

said to be constitutive of the presence of the 'cause' (Kripke 1982, 93fn). To take one of

Kripke's examples of a view reached by this means: "We do not condemn certain acts because

they are immoral; they are immoral because we condemn them" (93fn). Euthyphro's fallacy

might seem a particular risk in the context of this maneuver. But historically, reductionists have

rarely gone in for it. Thus the moral subjectivist of Kripke's example acknowledges that the

ordinary view and its "inverse" are in conflict; she does not hold, à la Euthyphro, that they can be correct together.

Contemporary naturalism pursues its reductive ambitions with a sophisticated conceptual arsenal that was unavailable to earlier reductionists. One danger of sophistication is that it can serve to paper over underlying confusions and inconsistencies, which in the context of the more austere formulations of past reductionists might have been exposed to the light of day. I will argue that this is just what has happened in the case of *informational semantics*, a central player in the contemporary literature on 'naturalizing' the mind. Peel away informational semantics' modern-day materialist trappings and we find one of the most primordial of philosophical confusions, Euthyphro's fallacy.

This result is of interest not simply in light of informational semantics' prominence in the literature. The argument to come assembles an array of considerations concerning the logic of reductive and naturalistic explanation, considerations whose implications for the naturalistic project in the philosophy of mind are potentially far-reaching. Criticisms of attempts to naturalize aspects of the mental typically focus on the criterion of extensional adequacy, raising doubts, via artfully constructed counterexamples, about the necessity or sufficiency of whatever account is under consideration. The drawback to this kind of objection is that its impact will be localized; a counterexample to one account will likely not serve as a counterexample to a marginally different proposal, and it will seem open to view the difficulty as calling for fine-tuning rather than abandonment of the account's core idea. One aim of this paper is to show that attention to more abstract considerations about the nature of explanation, especially about the interaction between naturalistic explanations and the familiar causal explanations of everyday thought, will reveal problems that cut deeper. Indeed, the constraints these considerations

impose on naturalistic accounts of the mental are severe enough to provoke a serious doubt about

the possibility of successful execution of the naturalistic project.

1. Informational semantics: a brief primer

Informational semantics is a view about the intentional contents of mental states and

occurrences: its core idea is that the contents of items in the head are constituted by the nomic

(i.e., lawful) relations these items bear to elements in the surrounding world. As the

informational semanticist Jerry Fodor notes, it helps in thinking about this idea to have a "zeroth-

order, tinker-toy approximation to such a theory to play around with" (Fodor 1994, 4). Let us

say that occurrences of an inner state *reliably correlate* with a property or kind k iff the state

occurs, as a matter of natural law, only when instances of k are present in one's surroundings.

So if a state of your brain is triggered, as a matter of law, only when dogs are around,

occurrences of that state reliably correlate with the property of doghood. A simple version of

informational semantics would hold that an inner occurrence's having the content *dog* consists in

its reliably correlating with doghood.[2]

Given the close relationship between laws and counterfactuals, we can equally understand

the basic information-semantic idea to be that content is determined, in Fodor's phrase, by an

inner state's "subjunctive career" (Fodor 1990a, 58). It's not enough that a given occurrence of a

state happens to be caused by a dog; the occurrence has the content *dog* only if that state

wouldn't occur if a dog *weren't* present.

It's easy to see what many naturalistically minded philosophers find compelling about this

idea. We find striking, if circumstantial, evidence for it in the fact that our ascriptions of content

to states of measuring instruments, such as thermometers, appear controlled by our beliefs about

what those states lawfully depend upon. Our inclination to report, upon examining a thermometer, that the thermometer "says" it's such-and-such temperature outside seems closely connected to our awareness that the level of the mercury in the thermometer lawfully co-varies with the temperature. Further evidence of this sort can easily be adduced.

A second consideration in favor of informational semantics is that the head/world lawful dependencies out of which it seeks to construct mental content are not, at least on the face of it, a philosopher's fancy. It's an obvious fact of our lives, not a theory-driven conjecture of the philosopher, that a competent adult is a reliable detector of the comings and goings in her environment of a wide range of kinds of physical object. (Were it not so, reliabilism, informational semantics' close cousin in the epistemological sphere, would not be regarded as an anti-skeptical account of knowledge.) And it seems plausible that the pertinent notion of reliability can be cashed out in terms of the presence of nomic or counterfactual relations between our thoughts about k's and k's in the world. Contrast the situation of one of informational semantics' chief competitors, 'biosemantics'. Biosemantics aims to show that facts about the contents of structures in our brains reduce to facts about the evolutionary history of these structures (Millikan 1989/1991). Setting aside the question of whether content possession could sensibly be regarded as constituted by such facts, the claim that structures in the brain were selected in the ways biosemanticists describe is no less speculative than the hypotheses advanced by evolutionary psychologists about the selectional pressures producing various social and interpersonal phenomena. Thus the advocate of biosemantics has double the work cut out for her as does the informational semanticist: she must not only defend the intelligibility of reducing content possession to her favored class of naturalistic fact; she must first convince us that this class is not empty.

As with any philosophical theory of anything, informational semantics faces its share of acknowledged difficulties. By far the most widely discussed stems from the fact, obvious in itself, that human beings can be wrong about what's going on around them. It's one thing to say that we are reliable detectors of ordinary kinds of object k; it's quite another to say that we are *infallible* detectors of k's. Such is clearly not the case: even the most sophisticated canine expert could, in sufficiently foggy conditions, think a dog was present when what's really there is, say, a wolf. It seems to follow that to whatever degree the expert's *dog*-thoughts nomically correlate with instantiations of the property *being a dog*, they correlate more strongly with instantiations of the disjunctive property *being a dog or a wolf-in-foggy-conditions*.[3] The challenge is to keep our information-semantic account of content from incorrectly assigning the disjunctive content to these thoughts.

A second worry is that informational semantics embodies an oversimplified conception of human cognition—in particular, a conception of our cognitive life as entirely given over to tracking changes in environmental conditions. In effect, the worry is that informational semantics conceives a human being in her cognitive aspect as nothing more than a sophisticated thermometer. A thermometer's 'attention' is fixed always on its immediate current environment; to the extent that it can be said to entertain a content at all—say, the content *65°F*—it is only in the context of a 'thought' about the current state of the environment—say, *it's 65°F here now*. This limitation seems precisely what makes it plausible to see a thermometer's entertaining the content *65°F* as lawfully correlating with the actual holding of that condition. Needless to say, our content-entertaining capacities are not similarly tethered. We may entertain thoughts wholly unprompted by what is going on around us at the time, we may entertain structurally complex thoughts involving multiple predicative or otherwise sub-propositional contents, and we may

adopt an array of attitudes toward a given thought, including attitudes like desire and intention that exhibit what Searle calls a world-to-mind direction of fit. It is obscure on the face of it how these myriad ways of entertaining the content k could all be thought to reliably correlate with, or otherwise lawfully depend upon, k's in the world. The question is how to make such instances of content-entertaining susceptible to information-semantic treatment.

I mention these concerns by way of clearing the plate: I will not develop them further, nor discuss any of the ingenious proposals in the literature for bending the basic informational story to accommodate them (see, e.g., Stampe 1977, Dretske 1981/1999, Fodor 1990b, and Fodor 1999). The argument I will develop is targeted to the core information-semantic thesis that intentional content reduces, in whole or in part, to nomic dependency, and it applies to *any* proposed account of content that incorporates that thesis. It will aid exposition to take as our foil the simple version of informational semantics described above; once the argument is complete, it will be easy to show that it generalizes.

2.   Informational semantics and intentional mechanisms

In light of the summary of informational semantics just given, the question posed in the title of this paper may seem inapt. Euthyphro's fallacy requires *two* (inconsistent) claims, and informational semantics, understood simply as the thesis that head/world nomic relations constitute intentional content, could at best serve as one member of such a pair. Nonetheless, the accusation is not unwarranted. For as we shall see, the additional claim needed to generate the inconsistency is annexed to the main information-semantic thesis by a number of prominent informational semanticists. Moreover, their endorsement of this claim is not a mere quirk on their part; as I shall discuss in section 7, the claim is probably unavoidable given the theoretical

mindset that motivates informational semantics.

As we have noted, informational semantics seems right to presume the existence of nomic correlations between our thoughts and what they are about. The claim now at issue addresses the question of how we might expect to explain, to account for, their existence. Possibly some laws of nature are *basic*—that is, they cannot be explained in terms of anything else, and must be regarded as among the ultimate building blocks of the universe. But it would be absurd to suppose that laws connecting *k*-thoughts and k's belong to this category. It's surely not a ground-level law of the universe that, say, such-and-such neurophysiological state of my brain occurs only in the presence of dogs. When such a law obtains, there must be an answer to the question *why* it obtains. As Fodor puts it, "if informational semantics is right about the metaphysics of meaning, there must be mechanisms *in virtue of which* mental (-cum-neural) structures 'resonate' to *doghood* and *Tuesdayhood*" (Fodor 1998, 75). To specify the mechanisms "in virtue of which" a law obtains—elsewhere Fodor speaks in the same spirit of the mechanisms that "implement" or "sustain" a law—is not simply to restate the law; it is to describe a further constellation of facts about the world that explains why the law holds.[4]

Fodor's own suggestion—and this is the claim I will argue is incompatible with the main information-semantic thesis—is that *k*-thought/k nomic correlations are typically, if not invariably, implemented by *intentional* mechanisms, by mechanisms whose specification requires mention of intentional contents of states and occurrences of the person in question. Call this principle **IM** (for 'intentional mechanism'):

> **IM.** Nomic correlations between *k*-thoughts and k's are implemented by intentional mechanisms.

What kind of intentional state or occurrence might figure in a mechanism implementing a *k*-thought/k nomic correlation? The most obvious candidate is *what the thinker knows about k-*

hood. Here's a simple example from Fodor: "Someone may rig things so that a bell goes off when the dog shakes its head. If I know how things are rigged, hearing the bell may reliably cause me to think *dog*" (Fodor 1998, 77). For a less artificial example, consider the case of a birdwatcher. A skilled birdwatcher has a wealth of knowledge about various species of birds—about what they look and sound like, about what activities they engage in, about where they are to be found and where not, about what counts as evidence for their presence and what does not, and so on. She may know, for example, that the common grackle has a long and sharp beak, that it's coloration varies in certain definite ways by region, that it has a distinctively harsh song, that it roosts and migrates in mixed company with many other species, and so on. Given her possession of such knowledge, she is disposed to infer the presence of common grackles from a large range of observed phenomena, and to refrain from doing so in a large range of other cases (including many cases where those less knowledgeable would be misled). If one has a view of perceptual experience according to which possession of salient knowledge can shape the very content of a person's experiences, one will also be open to the possibility that the birdwatcher can sometimes just see (or hear), without having to draw any inferences, that a common grackle is present, in circumstances where a similarly placed layperson would not. The precise character of the relevant mechanisms needn't concern us. What matters for current purposes is just the evident plausibility of the thought that the birdwatcher's possession of such knowledge is well suited to explain why she is a reliable detector of common grackles. And so it explains a *common-grackle*-thought/common-grackle nomic correlation, if such obtains.[5]

   IM holds that intentional explanations of *k*-thought/k nomic correlations are available generally—that the mechanisms gestured at in these examples are the rule rather than the exception. A number of informational semanticists explicitly endorse, indeed argue, for this

thesis. They include Fodor (1998), Eric Margolis (1999), Fiona Cowie (1999) and Murat

Aydede and Güven Güzeldere (forthcoming).[6]

I won't pause now to consider the reasons why these informational semanticists take such an

interest in **IM**, though I will partially return to the question in section 7.[7] Of more immediate

relevance is that Fodor in particular anticipates that one might perceive an incompatibility

between informational semantics and **IM**, and he's concerned to combat that perception. He

misses the real source of the incompatibility. But it will help in understanding that source if we

first examine why Fodor thinks worries about incompatibility are misplaced.

Fodor foresees two potential worries about the compatibility of **IM** with informational

semantics. First, one might suspect that taking the mechanisms sustaining content-constituting

relations to involve contentful states and occurrences is tantamount to holding that the content of

a thought is determined by its relations to other content-bearing items. This is straightforwardly

incompatible with informational semantics, for informational semantics is *atomistic*. In

maintaining that an item can possess intentional content just in virtue of its standing in the

appropriate nomic relationship to an element in the external world, it denies that the item's

membership in a system of content-bearing items plays any essential or constitutive role in

determining its content.

According to Fodor, this objection betrays a serious confusion (a confusion he attributes to

Quine):

> It's a typically Quineian move to argue that since the semantical relations between,
> as it might be, 'proton's and protons are *theory mediated* (since, in particular,
> theoretical inferences mediate our applications of 'proton' to protons), it must be that
> *what one means by 'proton' is partly determined by the theories about protons that
> one endorses*…But Quine is not a good Skinnerian in holding this. A good
> Skinnerian says that what 'proton' means is determined *just by its functional relation
> to* (its causal covariance with) *protons*; given that this covariation holds, the
> theoretical inferences by which it's mediated are *semantically irrelevant* (Fodor

1990a, 56, emphasis his).

As Fodor sees it, Quine infers his brand of holism from the observation that content-constituting relations are "theory mediated"—they hold because of the subject's disposition to make particular theoretically informed inferences. But, says Fodor, to endorse this inference is to miss that an account of content in terms of nomic covariation will simply abstract from the beliefs and inferences in virtue of which the covariation is mediated. Possession of a particular theory may well be *responsible* for a nomic connection between protons and some state of one's brain, but on the causal-covariation (i.e., informational) account of meaning, it is the holding of the nomic connection, and not the theory's playing its mediating role, that *constitutes* occurrences of that state's having the content *proton*. Two mental items caught up in radically different theories could have the same content, so long as the theories conspire to sustain the same type of nomic relationship to the same kind of physical object.

The second worry Fodor considers concerns not semantic holism but naturalism. Here the thought is simply that appeal to intentional implementing mechanisms clashes with informational semantics' commitment to semantic naturalism. How could it be acceptable to hold both that head/world nomic relationships are sustained by content-involving mechanisms and that those relationships serve as the naturalistic material out of which content is constituted?[8]

Here again, Fodor finds the worry to be born of a conflation of distinct issues:

> What meaning is, is a metaphysical question to which, I'm supposing, informational semantics is the answer. The current question, by contrast, is about not metaphysics but engineering: how are certain lawful mind-world correlations (the ones that informational semantics says are content-constituting) achieved and sustained? Answers to this engineering question can unquestion-beggingly appeal to the operation of semantic and intentional mechanisms, since 'semantic' and 'intentional' are presumed to be independently defined (Fodor 1998, 78-79).

The metaphysical question is, "What are the relationships between internal items and external kinds that constitute content?" The engineering question is, "What mechanisms are

responsible for the holding of these relationships?" Semantic naturalism is a view about what constitutes content, and as such imposes constraints on how we are to answer the metaphysical question. In particular, it requires that in answering that question we make no use of semantic or intentional vocabulary. But semantic naturalism imposes no constraint whatsoever on how we are to answer the engineering question. It is mute on the propriety of using semantic or intentional vocabulary in this context, just as it is mute on the propriety of using such vocabulary in any context at all other than when we are trying to give a metaphysical account of how content is constituted. Of course, as thoroughgoing naturalists, we would be astonished if all mentalistic phenomena didn't ultimately prove tractable to naturalistic analysis. But even if that weren't the case, and so answers to the engineering question had to make ineliminable use of intentional terminology, this would not by itself undermine the claim of informational semantics to have naturalized intentional content, for it would not change the fact that informational semantics' answer to the metaphysical question can be put in naturalistic terms. The only way one might go astray here would be to explicitly link the two questions together in one's answer to the metaphysical question, by, say, identifying the content-constituting relations as those nomic relations that are sustained by mechanisms of a certain, intentionalistically described, sort. But informational semantics does not do this; as Fodor says, content on that view is "defined independently" of how we answer the engineering question.

3. Euthyphro's fallacy

Contrary to the claims of Fodor and his allies, there *is* an incompatibility between **IM** and informational semantics. But it's one we won't be in a position to see without bringing into view an important point about the logic of explanation.

Informational semantics, as with any form of semantic naturalism, makes what we can call a

constitutive claim. A constitutive claim about a property or kind, say the property of being an f, is a claim about *what it is to be an f*, about *what f-ness consists in*. Constitutive claims are ubiquitous inside philosophy and out, and may be defended on grounds of a widely divergent character. At one extreme, a claim about what it is to be an f may be based on considerations that are alleged to emerge simply through armchair reflection on the concept of an f. At the other extreme, and in particular when f's constitute a natural kind, a claim about what it is to be an f may be presented as the empirical deliverance of a natural science. Informational semantics, and other forms of semantic naturalism, are probably best understood as defended on grounds of a third sort—on what we might call 'naturalistic' grounds. Naturalistic arguments for constitutive claims share features with those of the other two categories. On the one hand, a naturalistic argument for a constitutive claim about f-ness will portray the claim as suggested by central features of our ordinary, pre-philosophical, pre-scientific understanding of f-ness. On the other hand, the argument will be shaped by a condition that will appear peripheral, at least at first glance, to our concept of an f—namely, the naturalist's constricted conception of the natural. The basic idea of a naturalistic argument for a constitutive claim about intentional content is that possession of content must consist in g, because g is the closest anything comes in the 'natural' world to having the features and playing the roles that we pre-philosophically take content to have and to play.[9]

Now, the point we need to take account of is the following. (Note that by "implemented law" in this formulation, I mean a non-basic law, a law implemented by a mechanism.)

> **S.** *Socrates' principle.* If it follows from a true constitutive claim that all f's are g's, then it's not an implemented law that all f's are g's.

If f-ness and g-ness are constitutively linked—if it is in virtue of what it is to be an f that an f is a g—there is no explaining the link between f-ness and g-ness in terms of an implementing

mechanism. The fact that f-ness insures g-ness cannot be thus overdetermined, explained both as an implication of what it is to be an f and as the result of an implementing mechanism.

Socrates' principle, once grasped, is nearly self-evident: to deny it could only bespeak a failure to understand what a constitutive claim is supposed to be. It is precisely the point of taking something to be a constitutive fact (or to be implied by such a fact) that an explanation in terms of an implementing mechanism is regarded as inappropriate, indeed, as senseless. Consider a famous example from Kripke (1980). The fact that a substance's having the atomic number of 79 insures that it will be malleable is not a basic law of the universe; its holding can be explained by appeal to underlying processes governed by more fundamental laws of molecular chemistry and physics. If, with Kripke, we suppose that to be gold simply *is* to be the substance with the atomic number of 79, the same mechanisms will also explain why a substance's being gold insures that it will be malleable. However, if we adopt the more traditional supposition that being gold consists in possessing a complex of manifest and dispositional features, among them malleability, this explanation loses its cogency and point. We still require an explanation of why a substance with the atomic number of 79 will be malleable. But however we answer this question, that answer will not be any part of an explanation of why gold is malleable. Being malleable is simply part of what it *is* to be gold. Thus there is neither need nor place for an explanation of the connection between gold-ness and malleability in terms of physical or chemical laws and processes.

This point, by the way, does not imply that the identification of gold with the substance having the atomic number of 79, or other constitutive claims about natural-kind properties, cannot be part of a scientific theory. It is one of the revolutionary implications of the Kripkean view of natural kinds that a wide and central class of constitutive facts is to be discovered

empirically, and so that scientists are in the business of advancing and defending claims about 'essences'. What this shows, if it is true, is that scientists in the special sciences engage in at least two different styles of explanation. They explain why implemented laws hold on the basis of mechanisms governed by other (and generally more fundamental) natural laws, and they explain what various natural-kind properties consist in. Presumably they proceed on both fronts simultaneously, tying together explanations of the two sorts in the course of developing a complete theory of their subject matter. The current point is only that the two sorts of explanation are mutually exclusive on a case-by-case basis. A given connection between properties may be said to hold either in virtue of an implementing mechanism or as an implication of the correct understanding of what constitutes a certain property, but it cannot sensibly be said to hold in virtue of both.

Socrates' principle holds regardless of the nature of the kinds or properties at issue. It holds even if the focus of the constitutive claim is an *external* property, a property whose holding of an object involves that object's standing in some relationship, such as a causal or nomic one, to other objects. Consider the property of being loved by the gods. As Socrates brought Euthyphro to realize, one cannot have it both ways.[10] One cannot say both that piety consists in being loved by the gods and that the gods love pious things because they are pious. The latter claim says that the correspondence between piety and god-belovedness is the upshot of a lawful or counterfactual connection between the two properties, sustained, presumably, by a capacity on the part of the gods to discern piety wherever it may be and a disposition to love it when they see it. But this picture is precisely what we rule out when we view the property of being pious as constituted by the property of being loved by the gods. Just as we cannot identify the property of being gold with the property of having the atomic number of 79 and suppose that a place still

remains for an explanation of their correspondence in terms of physical or chemical processes, so we cannot take god-belovedness to constitute piety and suppose that a place still remains for an explanation of their correspondence in terms of the gods' powers of discernment or some other intervening mechanism. There is no room, as it were, for such a mechanism to occupy, no gap for it to bridge.[11]

As I have said, I take it that **S**, once clearly understood, is close to self-evident. However, certain logical and conceptual confusions can stand in the way of a clear understanding of the content of the principle. Rather than holding up the exposition of the argument, I reserve discussion of what I take to be the main danger here until the end of the paper (section 8). Let us turn now to the implications of **S** for informational semantics.

According to the simple version of informational semantics, it's constitutive of possession of the content *dog* that the item in question reliably correlate with doghood. Now, it follows from the definition of reliable correlation given in section 1 that an item that reliably correlates with doghood occurs in the presence of dogs. So according to informational semantics, it follows from what it is to have the content *dog*, from a constitutive claim about the property of having the content *dog*, that all items with the content *dog* occur in the presence of dogs. Given **S**, then, we can't endorse informational semantics and also suppose it's an implemented law, a law sustained by an implementing mechanism, that items with the content *dog* occur in the presence of dogs. If the property of *being an item with the content dog* and the property of *occurring in the presence of dogs* are constitutively linked—linked in virtue of what it is to have the content *dog*—it must be wrong to suppose that their correspondence is sustained by an implementing mechanism.

Let us call this conclusion **C1**:

C1.   If informational semantics is true, it's not an implemented law that all *k*-thoughts occur in the presence of k's.

4.   Socrates' principle, informational semantics, and the representational view of the mind

If we're not careful, it might seem that **C1** by itself is inconsistent with informational semantics, and it's important to see why this is not so.

Why might **C1** by itself seem to conflict with informational semantics? Well, as we've noted, laws concerning the relationships between items in the head and middle-sized kinds in the world are not basic laws of the universe. So whenever *k*-thoughts reliably correlate with k-hood in the world, as informational semantics presumes they will, the law underwriting the correlation will be sustained by an implementing mechanism. But according to **C1**, if informational semantics is true, it *can't* be an implemented law that *k*-thoughts occur in the presence of k's. And so we seem to have a problem.

But there is in fact no conflict here. What this argument misses is that sentences of the form, "f's reliably correlate with k-hood," are extensional: their truth-values are determined only by the extensions of the expressions they contain. To reliably correlate with k-hood, recall, is to belong to a type (for example, occurrences of a given inner state) such that it is a law that items of that type occur in the presence of k's. This definition does contain the operator, "it is a law that," but that operator's intensionality does not infect talk of reliable correlation itself. The reason is this: although an item may be said to belong to a given type in virtue of one of its properties and not in virtue of others, it would be a serious confusion, akin to the sort Davidson famously diagnosed, to conclude that it belongs to that type *only under certain descriptions*.[12] It is in virtue of his having a certain gene, and not in virtue of his current spatial location, that the man

16

in the next room is subsumed by a law that puts him at a high risk for cancer. But it is no less

true that the man in the next room, *described as such*, is at a high risk for cancer.

The current point is exactly analogous. In order for f's to reliably correlate with k-hood, it's

not necessary that it be a law that all f's occur in the presence of k's. It suffices that there be

some predicate, "is an h," such that everything that is an f is an h, and it's a law that all h's occur

in the presence of k's.[13] For example, suppose it's a law that elephant stampedes occur only in

the presence of tigers. And suppose, as it happens, that all the interesting events I saw last week

were elephant stampedes. Then all the interesting events I saw last week reliably correlated with

tigerhood. But of course it's not itself a law of nature that all the interesting events I saw last

week occurred in the presence of tigers. The predicate, "is an interesting event I saw last week,"

isn't the kind of predicate we'd expect to figure in a statement of natural law, and nothing in the

example requires that it do so. The reason all the interesting events I saw last week reliably

correlated with tigerhood is that these events were, in point of fact, elephant stampedes, and it's a

law that all *elephant stampedes* occur in the presence of tigers.

The same goes for informational semantics. According to **C1**, the informational semanticist

cannot suppose that it is an implemented law that all items with the content *dog* occur in the

presence of dogs. But for all we've said so far, she doesn't have to. All we need is that there be

some predicate, "is an h," such that items with the content *dog* are h's, and it's an implemented

law that all h's occur in the presence of dogs. If these conditions are fulfilled, then items with

the content *dog* reliably correlate with doghood without its being an implemented law that items

with the content *dog* occur in the presence of dogs.

And there's no special difficulty in finding potential candidates to replace "is an h" in

sentences like those above. All one needs is to endorse to the familiar view—what we might call

the "representational view of the mind"—that mental content is borne by physical items, items individuated by their membership in one or another physical kind. Most commonly, the relevant kinds are said to be either neurophysiological or functional. Fodor, for example, proposes that the vehicles of mental content are tokenings of expressions in 'mentalese', in a language of thought with a functionally individuated syntax.[14] If we adopt Fodor's view, we have no difficulty in finding predicates to replace "is an h" in sentences like those above. The appropriate predicates are predicates of mentalese syntactic types. The idea is that there will be a mentalese expression, say, "DOG", such that it is a law that tokenings of "DOG" in my brain occur only in the presence of dogs. Thus "DOG" tokenings will reliably correlate with doghood. In virtue of this fact, according to the simple version of informational semantics, tokenings of "DOG" will count as having the content *dog*. And so items with the content *dog* will reliably correlate with doghood. But it will not itself be an implemented law that items with the content *dog* occur only in the presence of dogs. The implemented law that underwrites the reliable correlation will be, rather, that tokenings of "DOG" occur only in the presence of dogs. And so we avoid conflict with **C1**.

5. Intentional mechanisms and cross-level implementation

Socrates' principle cannot by itself produce a difficulty for informational semantics. Nor, as we saw earlier, can we impugn the view solely on the basis of the fact (what I called **IM**) that thought/world reliable correlations are sustained by mechanisms involving the contents of states and occurrences on the part of the thinker. The trouble for informational semantics arises when we combine these considerations. To make the connection, we will need to venture one further premise (in a failure of imagination, I'll just call it "**P"** for "premise"):

P. If *k*-thoughts reliably correlate with k-hood, and if this correlation is implemented by an intentional mechanism, then it is an implemented law that all *k*-thoughts occur in the presence of k's.

One way to think of **P** is as an instance of a general truism about what in the philosophy of science is sometimes called "cross-level" implementation. It's common to speak of laws and mechanisms as organized into levels—the chemical level, the neurophysiological level, the intentional level, and so on. What places a law or mechanism at the chemical level is that its expression requires chemical vocabulary. Now, there is no problem with the idea of a mechanism at one level sustaining a law at a different level; much scientific endeavor is devoted to explicating phenomena couched in the vocabulary of one level in terms of mechanisms operative at a more fundamental level—for example, explaining macroscopic laws in terms of microscopic processes. But for a cross-level implementation to occur, the candidate implementing mechanism must first sustain a law at its own level. Unless a microphysical process *directly* implements a microphysical law, it cannot *indirectly* implement a macrophysical law.

For example, suppose the macrophysical law that all M_1's cause M_2's is implemented by a microphysical process in which an event of type m_1 initiates a chain of events culminating in an event of type m_2. For this to be so, it is necessary that the property of being an M_1 and the property of being an m_1 be either nomically or constitutively linked such that an occurrence of an M_1 suffices for an occurrence of an m_1, and that the property of being an M_2 and the property of being an m_2 be either constitutively or nomically linked such that the occurrence of an m_2 suffices for the occurrence of an M_2. Secondly, and this is the crucial point, it is necessary that the implementing mechanism sustain a law at its own level. The connection between M_1 and m_1, and that between M_2 and m_2, will obviously not help to explain why the macrophysical law holds

unless it is itself a law that all m_1's cause m_2's. Unless the microphysical process directly

implements the law that all m_1's cause m_2's, it cannot indirectly implement the law that all M_1's

cause M_2's.

Precisely the same goes for the intentional mechanisms that sustain head/world laws.

Intentional mechanisms occupy the intentional level, the level of mechanisms and laws whose

specification requires mention of the intentional contents of beliefs or other content-bearing

states and occurrences. Thus our truism about cross-level implementation implies that, if an

intentional mechanism is to implement a law at any level, it must implement an intentional law.

Since the intentional mechanisms we are envisioning are geared in particular toward sustaining

laws that underwrite *k*-thought/k reliable correlations, they must sustain laws of the form, "All *k*-

thoughts occur in the presence of k's." That is our **P**.

It is worth emphasizing that there is no implication here that intentional mechanisms cannot

sustain non-intentional laws, such as neurophysiological or computational laws. The claim is

just a conditional one: if an intentional mechanism is to sustain any law at all, it must sustain an

intentional law.

6. The incompatibility of informational semantics and IM

All the pieces are in place. The argument can now be made very quickly.

We have already discussed how **S** yields **C1**. According to **S**, if it follows from a true

constitutive claim that all f's are g's, it cannot be an implemented law that all f's are g's. The

correspondence between these properties cannot be overdetermined in this way, explained both

by what it is to be an f and by an intervening causal process or other mechanism. And so if

informational semantics is right that it follows from the correct account of what constitutes

possession of the mental content *k* that all thoughts with the content *k* occur in the presence of

k's, then that correlation can't also be explained by an implementing mechanism. And that's **C1**:

> **C1.** If informational semantics is true, it is not an implemented law that all *k*-thoughts occur in the presence of k's.

That completes the first part of the argument.

Now for the second. According to informational semantics, *k*-thoughts reliably correlate

with k-hood. According to **IM**, if a person's *k*-thoughts reliably correlate with k-hood, this

correlation is sustained by an intentional mechanism. And according to **P**, if *k*-thoughts reliably

correlate with k-hood, and if this correlation is sustained by an intentional mechanism, then it is

an implemented law that all *k*-thoughts occur in the presence of k's. Thus if informational

semantics is right, it *is* an implemented law that all *k*-thoughts occur in the presence of k's. Call

that **C2**:

> **C2.** If informational semantics is true, it is an implemented law that all *k*-thoughts occur in the presence of k's.

But if **C1** and **C2** are both true, then supposing that informational semantics is true produces

a contradiction. It cannot both be an implemented law, and not be an implemented law, that all

items with the content *k* occur in the presence of k's. I've argued that **S** and **P** are undeniable.

Short of abandoning **IM**—short, that is, of abandoning the idea that it is a person's knowledge of

the environment around her that explains why she's reliable about what comes and goes within

it—we must abandon our supposition that informational semantics is true.

How did Fodor and his peers end up committing a mistake as fundamental as Euthyphro's

fallacy, and why has that mistake escaped notice? The answer, I think, is that the mistake

emerges only when we combine considerations that are generally not scrutinized together.

Looking at just **IM** in isolation, it will seem that Fodor's distinction between the metaphysical and engineering questions suffices to ward off any threat of incompatibility between **IM** and informational semantics. Informational semanticists can accept that intentional mechanisms explain thought/world reliable correlations without violating naturalism about mental content, because the question of what *constitutes* content is one thing, and the question of what *sustains* content-constituting relations is another. And on the other hand, considering just Socrates' principle (**S**) in isolation, it will seem that the representational view of the mind is enough to keep informational semantics from running into conflict with the principle. Socrates' principle entails that we cannot suppose both that having the content *k* consists in reliably correlating with k-hood, and that it's an implemented law that *k*-thoughts occur in the presence of k's. But given the representational view of the mind, it would seem that the informational semanticist doesn't have to posit the existence of such laws. Instead, she can take reliable correlations between *k*-thoughts and k's to be underwritten by laws couched at a physical level, say that of neurophysiology or of mentalese syntax.

It's when we take **IM** and Socrates' principle *together* that we get in trouble. For, as we have seen, accepting that thought/world correlations are sustained by intentional mechanisms requires us also to accept that the implemented laws underwriting these correlations are at the intentional level. And so the appeal to the representational view of the mind cannot help us. Accepting **IM,** the informational semanticist must accept that it's an implemented law that *k*-thoughts occur in the presence of k's. But given Socrates' principle, an informational semanticist *cannot* accept the existence of that implemented law, on pain of committing Euthyphro's fallacy.

As I noted at the outset, I have built the argument around a simple version of informational

semantics. This allowed for a more transparent presentation of the problem. But it should be obvious that the argument generalizes to any version of informational semantics—that is, to any account of content that sees content as constituted, partly or wholly, by thought/world nomic relations. There are three ways in which an informational account of content might differ from the simple version considered here: 1) by appealing to nomic relationships other than that of strict reliable correlation—for example by using a weaker form of lawful dependency or by requiring 'world-head' rather than (or in addition to) 'head-world' reliability,[15] 2) by restricting the class of inner items whose lawful relationships matter for the determination of content—for example by restricting the class to those occurrences of inner states that satisfy a naturalistically specified analogue of the idea of "observationally ideal" circumstances for identification of external kinds,[16] or 3) by casting an inner item's standing in a nomic relationship to k-ness as just one element in a package of conditions that must be satisfied if that item is to possess the content *k*—for example Fodor's wedding of the basic informational condition to a complicated requirement involving something he calls "asymmetric dependency".[17] It should be clear that no fine-tuning of the first two varieties will have any bearing on the issue raised in this paper. The relevant point is that changes in the nature or scope of the nomic correlation that is said to obtain between *k*-thoughts and k's will not affect the plausibility of the claim that the correlation is implemented intentionally. For example, intentional mechanisms are equally fit to explain reliable correlation in the world-thought direction as in the reverse: what I know about dogs helps explain why I don't mistake a dog for something else no less than it helps explain why I don't mistake something else for a dog. A parallel of the argument given here is then easily constructed.

And, of course, modifications of the third sort, in which *further* conditions are tacked on to

the basic informational story, are irrelevant to the current difficulty. If it is our view that an inner

item's having the content *k* consists in its reliably correlating with the presence of k's plus its

satisfying some further condition X, then according to our view, it follows from a true

constitutive claim that *k*-thoughts reliably correlate with k-hood. **S** applies as before.

7. Why informational semanticists are more or less stuck with IM

This completes the argument that informational semantics is incompatible with **IM**. Those

informational semanticists who endorse **IM** commit Euthyphro's fallacy.

Establishing this incompatibility was the primary business of this paper. I believe that the

argument just presented for this result is of interest for metaphysics and the naturalistic project in

the philosophy of mind generally, and is not solely tied to our concern, such as it may be, with

the prospects for informational semantics. But insofar as we *are* concerned with informational

semantics' prospects, there is one more question we will want to address. Did Fodor et. al make

a gratuitous error in endorsing **IM**? These informational semanticists seem to think they are

committed to **IM**.[18] But knowing what we know now, why can't we just recant this view on

their behalf?

The first thing to say on this question is that to reject **IM** across the board—to claim that *k*-

thought/k nomic correlations are *never* intentionally implemented—would put one in serious

conflict with our ordinary, 'folk'-psychological thought. It is simply a fact about our everyday

explanatory practice that, in at least a wide range of cases, we explain a person's capacity to

reliably identify members of a kind—and so to have her thoughts about that kind reliably

correlate with its local exemplifications—in terms of what she knows or believes about that kind.

Recall our birdwatcher. Why does the birdwatcher have the ability to reliably identify common

grackles as such when I do not? Surely the answer the folk would give to this question is that the birdwatcher knows many things about the common grackle—about its appearance, habits and milieu—that I do not. For the more practically minded sort of birdwatcher, the whole *point* of accruing knowledge about a given bird species is that such knowledge enables reliable identifications of members of that species. If she didn't believe that knowledge could place this role, she wouldn't waste her time studying field guides. Other examples of this sort can easily be assembled. It doesn't take much reflection to realize that it is a pervasive assumption on the part of us folk that the accuracy and reliability of a person's identifications of kinds can be explained by what she knows or believes about those kinds.

Perhaps we folk are wrong on this score, but why should one think so? The only principled basis I can see for categorically denying these explanations would be a still more categorical denial of the legitimacy of folk-psychological explanation as such. If the very strategy of citing the contents of a person's propositional attitudes to explain things about that person is wrongheaded, then of course the particular applications of that strategy now in question are mistaken too. A number of philosophers have in fact argued for this dismissive view of folk psychology. Whatever the merits of their arguments, all that matters for our purposes is that this tack is not open to an informational semanticist. Informational semantics is a form of semantic naturalism: its aim is to show that facts about possession of intentional content reduce to facts expressible in natural-scientific (hence non-intentional and non-semantic) terms. As such, its motivating conviction is precisely *not* that commonsense intentional psychology is fundamentally incorrect or confused. Semantic naturalists are not eliminativists; their aim is not to abandon folk psychology but to *vindicate* it by showing the place of the notion of intentional content in a naturalistic conception of the world. Informational semanticists like Fodor see

themselves as folk psychology's champions. Thus they are not liberty to dismiss its basic

presupposition of the explanatory relevance of intentional content. And if they are not at liberty

to dismiss that, they are not at liberty to dismiss one of its more characteristic employments:

explaining the reliability of a person's beliefs about what is present in her surroundings in terms

of her knowledge and beliefs about the relevant aspects of the world.

A more moderate thought, one that might seem open to the informational semanticist even

given her commitment to folk psychology, is that we don't have to accept **IM** in its full

generality. As we turn our attention away from the fine-grained and sophisticated identifications

of the birdwatcher to the layperson's identification of more familiar kinds and properties, the role

of knowledge in sustaining reliability might start to seem less obvious. The possibility may seem

to arise that there is a core set of kinds k such that *k*-thought/k nomic correlations are not

mediated at all by any knowledge on the part of the thinker, and so may be presumed to be

sustained by wholly non-intentional mechanisms. Cowie, for one, thinks that this is so: in the

case of certain simple kinds, she suggests, our minds "resonate" to these kinds solely in virtue of

unlearned and non-intentional "cognitive reflexes" (Cowie 1999, 134ff).[19] If she is right, then

maybe we can partially salvage informational semantics by restricting its scope to these cases.

This is not Cowie's own view, of course: she holds with Fodor and the others that informational

semantics has a general field of application. But if we suppose instead that content is constituted

by head/world nomic relations only in those instances where the relations are not implemented

by intentional mechanisms, we avoid conflict with **IM**.[20]

A fully satisfactory answer to this question—the question being whether, and to what extent,

there are *k*-thought/k nomic correlations none of whose sustaining mechanisms are knowledge-

involving or otherwise intentional—would require inquiring in detail into the nature, structure

and variety of explanations in folk psychology. It would require an account of what intentional mechanisms actually *look like*—in their interiors, so to speak—a topic we did not need to broach for the argument above and which I do not have the space to adequately address now. But I will briefly touch on a few considerations that suggest that Cowie's vision of a core of non-intentionally implemented head/world nomic correlations is much less plausible than it might seem at first glance.

First, it's important to be clear on exactly what Cowie's view requires. The claim is not merely that non-intentional mechanisms sometimes have a hand in implementing *k*-thought/k nomic correlations. *That* is certainly a reasonable thought: it's undeniable, after all, that there are neurophysiological and computational (or otherwise cognitive-scientific) mechanisms operative in perception, and detection of external kinds obviously involves perception. But merely positing, however plausibly, the presence of such a mechanism in a given instance of a *k*-thought/k nomic correlation does not imply, or even suggest, that intentional mechanisms are not present as well. As Fodor has long argued (1980 and 1997), natural laws can be—and in the case of the laws of many special sciences, generally are—*multiply* implemented, sustained by a mix of different mechanisms. What the current proposal requires, however, is not just that there are *k*-thought/k nomic correlations with non-intentional implementing mechanisms, but that there are *k*-thought/k nomic correlations without any intentional implementing mechanisms. For as long as one of the mechanisms linking *k*-thoughts to k's is intentional, premise **P** applies, and the rest of the argument falls into place.

Second, the view ought not to accrue unwarranted credibility from being mistakenly viewed as a consequence of the idea that some concepts are innate. Cowie is moved by the "classical empiricist" thought that an initial set of concepts must be unlearned if any are to be learnable.

Given her information-semantic view of concept possession, that thought comes to this: "there are some properties we are born able to detect," and "it is our ability to resonate with these basic properties that presumably enables us to come to detect the rest" (Cowie 1999, 133). Cowie suggests that in at least some cases, the mechanisms enabling our innate detecting capacities will be "cognitive reflexes", non-psychologically mediated dispositions to enter into given brain states in the presence of the kinds or properties in question. This thought in turn motivates a qualification to her claim (a version of **IM**) that "given an informational account of concepts, possessing a concept…will involve having knowledge of some kind about its object," the qualification being: "The only exception is concepts that we possess in virtue of our 'cognitive reflexes'" (Cowie 1999, 135).

But who is the "we" Cowie is talking about here? Surely adult human beings, the full-fledged thinkers and subjects that any account of mental content must take as its primary concern. Thus Cowie has made an invalid inference: she moves directly from the premise that certain capacities are wholly non-intentionally implemented *at birth* to the conclusion that those capacities are wholly non-intentionally implemented *at maturity*. This conclusion is justified only given the further premise, unstated by Cowie, that no intentional mechanisms come to supplement the original "cognitive reflexes" at any stage in the maturity of the subject.

Once this additional premise is brought out into the open, it seems to me it ought not to survive scrutiny. There is a salient difference between the infant that detects certain "basic" properties without relying on any knowledge of them and the adult that Cowie envisions as doing the same: namely, the adult *has* knowledge about these properties. Adults know things—many, many things—about colors, shapes and whatever other properties might strike one as "basic"; acquiring such knowledge, after all, is part of what maturation is. Since they have such

knowledge, and since such knowledge is surely at least sometimes germane to the identification of instances of the properties in question, on what ground can we have confidence that the knowledge will not end up factoring in any mechanisms sustaining the relevant correlations? Of course, if infants have the capacity to reliably identify certain properties in the absence of any knowledge about them, then it's logically possible that adults do not rely on their knowledge either. But our concern is not with what is possible: it is with how thought/world correlations are actually sustained in the case of real live adult human beings. The reasonable thought is surely that they will use whatever tools they have at their disposal, knowledge being one such tool, and a prominent one at that.

These considerations tend toward the conclusion that we can avoid conflict between informational semantics and **IM** only by restricting the scope of the information-semantic account to the case of simple concepts *when possessed by children or simple animals*. As soon as a child reaches a developmental stage at which intentional mechanisms emerge to supplement innately given non-intentional "reflexes", the argument of this paper shows that concept possession can no longer be understood in information-semantic terms. The role this leaves for informational semantics seems too residual (not to mention *ad hoc*) to have much appeal.

Finally, one last way to try to avoid conflict with **IM** is to recast informational semantics as an account of subpersonal content. So long as one then resists the temptation to suppose that subpersonal psychological happenings constitute personal-level mental life, there will be no conflict with **IM**. But by the same token, the prospect of this recasting is irrelevant to our topic, since informational semantics so conceived is not an attempt to realize semantic naturalism. In any case, it's likely that an analogue for **IM** holds for subpersonal psychology with complete generality, at least if we assume anything like a computational framework. As Cummins notes,

"The CTC [computational theory of cognition] has it (indeed, this was the fundamental claim of the so-called cognitive revolution) that cognitive systems are able to get into states that reliably covary with distal features because of their stored knowledge…A computational system of the sort favored by the CTC has no hope of arriving at the truth about even very common *perceptual* matters without the help of a formidable background of *knowledge [sic]" (1989, 44-45).

8.   A final word on Socrates' principle: constitutive vs. logical relations

Finally, the piece of unfinished business left over from section 3.

The primary obstacle to a clear understanding of **S** is presented by a familiar and, at least in the philosophy of mind, perennially tempting category mistake: the conflation of a thing with its description. This mistake bedevils discussion on the closely related topic of the "causal relevance" of dispositional and functional properties, and it will be helpful, before concluding, to briefly consider how the difficulty can shade over to obscure one's view of **S** as well.

In his well-known discussion of the question whether second-order (functional) properties can be "nomologically sufficient" for "the effects in terms of which they are defined" (and hence, on the "nomist conception of causal relevance", causally relevant to these effects), Ned Block writes:

> 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*…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…. [Properties can be] logically related under one set of descriptions, nomologically under another (1990, 157-158).

The bearing of these remarks on our topic at first seems clear. By "logically related properties", Block appears to mean something akin to what I would prefer to call "constitutively related properties", that is, properties whose correlating in some particular way is guaranteed by facts about what possession of one or the other of the properties consists in. Thus Block's initial remarks in the passage suggest he inclines toward a principle very similar to, and in fact stronger than, **S**, namely, that two properties cannot be both constitutively and nomically related. (In less ontologically committed terms: if it follows from what it is to be an f that all f's are g's, it's not a law that all f's are g's.) However, in the second part of the quoted passage, Block indicates that he doesn't wish to assert any such principle, that to do so, in fact, would be to commit a "fallacy". His final, considered position is that properties *can* be both nomically and "logically" related (although, he adds, it would be "amazing" if this were the usual case) (1990, 158).

But what "fallacy" is supposed to be at work here, and what is its analogue for "Davidsonian token events"? Presumably Block is thinking of Davidson's famous response to A.I. Melden's view that a cause must be "logically distinct from the alleged effect" (Quoted at Davidson 1963/1980, 13). In opposing this view, Davidson points out that the members of any cause-effect pair can be given logically related descriptions: e.g., "The cause of B caused B."[21] It is evidently this sort of Davidsonian consideration that is meant to fund Block's talk of properties being "logically related under one set of descriptions, nomologically related under another".

But in fact the relevance of the Davidsonian material to Block's assertions is quite equivocal. Davidson is speaking of "logical relations" in the ordinary sense of that phrase; that is, he is speaking of semantical relations that hold between linguistic items, for example, descriptions of events or properties, not relations between events or properties themselves. Facts about the possibility of various ways of describing events or properties are one thing; facts about the

metaphysical character of causal, nomic or constitutive relations between events or properties are quite another. For example, we can, on analogy to Davidson, give any nomically related pair of properties logically related descriptions: e.g., "The property that nomically necessitates B-ness nomically necessitates B-ness." But surely we would not want to conclude on this basis that all nomically related properties are also constitutively related. The availability of logically related descriptions of properties simply has no bearing on questions about the nature or extent of constitutive relations between properties.

For Davidson, the real problem with a view like Melden's is that it confuses relations between descriptions with relations between the items described: "In any case there is something very odd in the idea that causal relations are empirical rather than logical. What can this mean?…The truth of a causal statement depends on *what* events are described; its status as analytic or synthetic depends on *how* the events are described" (1963/1980, 14). Block, who couples talk of "logically related properties" with an appeal to Davidson's point about logically related descriptions, would appear guilty of precisely this mistake. He contrives to dissent from **S** (or a near cousin of **S**) only by conflating the idea of a constitutive connection between properties with the idea of a logical relation between descriptions of properties.

9. Conclusion

What we come to believe about the surrounding world is partly governed by what we already believe about the surrounding world—this much 'belief holism' is indisputable, or, at any rate, indisputably a part of our commonsense psychology. In particular, if a person's thoughts about the comings and goings of certain objects correspond in a lawful way to the actual whereabouts of those objects, we will find at least part of the explanation for this phenomenon in what the

person knows and believes about those objects and their environment. Fodor and other informational semanticists find no obstacle to their embracing this folk-psychological platitude. They see no inconsistency in supposing both that lawful relationships to external kinds constitute content, and that lawful relationships between kinds and thoughts are sustained by the subject's beliefs and other contentful states. I have argued that the informational semanticists are mistaken. There *is* an inconsistency between these suppositions. The result is that informational semantics and commonsense psychology cannot be integrated; one of them must go. Since informational semantics cannot sensibly do without commonsense psychology, there is no real choice to be had.

I want to conclude by suggesting that we have reason to suspect that informational semantics' troubles, as developed here, are representative of a problem for the semantic-naturalistic project as such. A naturalistic account of content must satisfy two requirements: it must portray possession of intentional content as consisting in the possession of non-mentalistically specifiable properties, and it must not conflict in any radical way with our ordinary, pre-philosophical intentional psychology. No semantic naturalist would dispute the first condition; it is definitive of her enterprise. Nor, as we saw earlier, would any semantic naturalist dispute the second.

One intent of this discussion is to sow a seed of suspicion that there is something delusive in the very idea that an account of content could jointly honor these two commitments. In casting about for material to help discharge the reductive requirement, a semantic naturalist will look to non-mentalistically describable properties that contentful states and occurrences can manifestly be seen to have, and perhaps also to features credited to these states and occurrences by a biological or cognitive-scientific theory. The problem is not that ordinary psychology denies that

contentful mental items have these features, or that it is inconsistent with ordinary psychology to suppose that they do. The threat of inconsistency arises rather when we take the step of supposing that the content of mental items reduces to some selection of these features. For in doing so, we cast as *constitutive* of content features and relationships that we were pre-philosophically disposed to regard as at best *accompaniments* of content. Surely it should come as no surprise if, in reorganizing the conceptual landscape so fundamentally, we cannot help but upset some significant portion of our ordinary beliefs about the nature of minds. And if that is what happens, we fail to honor our second requirement, the requirement to leave ordinary psychology largely as is.

Left in the abstract, this worry is obviously inconclusive, and that may help explain why semantic naturalists have in general shown little sensitivity to the possibility of a danger of this kind. This paper is an attempt to show in detail how the worry plays out in one concrete, and prominent, case. Whether other forms of semantic naturalism are afflicted with similar problems is an open question. But our current result shows that it is certainly a live one.[22]

---

[1] I shall expand upon and modify this formulation later on.

[2] A few historical points bear brief mention here. **1.** In his seminal account of informational semantics (1981/1999), Fred Dretske presents content as reducing to *probabilistic* rather than *nomic* head/world relationships. I follow later informational semanticists in opting instead for the latter. Some balk at the very idea of natural laws of the kind informational semanticists are thus led to posit, but I set such worries aside here. **2.** Note that it is occurrences of a state, and not the state itself (except derivatively), that are defined as "reliably correlating" with a kind or property. Here I follow Dretske (1990, 820) in casting tokens, rather than types, as the items that

stand in content-constituting relationships with external kinds. (Of course, it is in virtue of its belonging to a particular type that a token reliably correlates with what it does.) Commentators on Dretske frequently get this wrong; see, for example, Godfrey-Smith 1992. **3.** Dretske himself is no longer an informational semanticist, but, as it is sometimes called, a 'teleofunctionalist'. According to Dretske's teleofunctional semantics, an inner item's possession of its content reduces to facts about the lawful relationships that items of that kind are *supposed* to bear, whether or not they do in fact. I believe this view is subject to an objection similar to the one I develop in this paper. See my "Teleofunctionalism and Psychological Explanation," forthcoming in *Pacific Philosophical Quarterly.*

[3] Terminological note: following Fodor, I use "*dog*-thought" as shorthand for "inner item with the content *dog*".

[4] For talk of "implementing mechanisms", see Fodor 1994, 8ff. A word on this terminology: in the philosophy of science, talk of "mechanisms" has come to have a special significance over the last few decades owing to its association with the "causal-mechanical" view of scientific explanation, a view intended to compete with Hempel's nomological-deductive account. In that context, it is a term of art whose import one is required to spell out with care. (For a recent example of work in this vein, with reference to others, see Glennan 2002.) Fodor's talk of "mechanisms" is not intended as a contribution to this debate; hence his failure to give a precise, technical meaning to the term is not a lacuna. He's no more beholden to provide such an account than is just anyone who speaks of causation obligated to offer an analysis of that concept.

[5] Again, it is worth pointing out what one does *not* need to discuss in this context. There are a number of contentious philosophical questions about how intentional explanations work— about how citing the contents of a person's knowledge, beliefs, and other states and attitudes can

explain what she does or thinks. It is hotly debated whether such explanations are in some distinctive sense normative, teleological, subjective, holistic, etc. We need not address these questions here (although see section 7). The examples in the text are just meant to serve as reminders of the kind of intentional explanations we might ordinarily give of thought/world nomic correlations, and it is not a condition on recognizing the intelligibility or plausibility of such specimens of intentional explanation that one have a comprehensive philosophical theory of that form of explanation. (Were it otherwise, day-to-day life would be a lot more difficult.) Something similar goes for the examples' unexamined employment of the philosophically vexed notion of knowledge. The only feature of these and comparable explanations that bears on the current topic is that they are intentional: what matters is not that the explanations appeal to one's *knowledge* that such-and-such but that they appeal to one's knowledge *that such-and-such*. Thus it is of no significance for our discussion if, convinced on philosophical grounds that knowledge is not a "pure" mental state, we seek to reinterpret our ordinary explanatory appeals to knowledge in terms of ascriptions of beliefs plus some "external" conditions. The involvement of beliefs insures that the mechanisms descried in these explanations are still intentional. (But for compelling arguments that ascriptions of knowledge do play an essential and indecomposable role in intentional explanation, see the first three chapters of Williamson, 2000.)

[6] It's especially noteworthy that Cowie, whose book is largely an unremitting attack on Fodorian philosophy of mind, concurs with Fodor in endorsing both informational semantics and **IM**.

[7] One source of their interest that I won't discuss is their belief that the marriage of informational semantics and **IM** opens the door to novel solutions to various problems in the philosophy of mind.

[8] For a version of this objection, see Cummins 1989, 62-66.

[9] There are many questions that could be asked about the nature and rationale of the 'naturalistic' way of defending a constitutive claim. For an illuminating attempt to say what naturalists are up to that accords with the description given in the text, see Dretske 1994. For an attempt to undermine one allegedly central rationale for the naturalistic enterprise, see McDowell 1994 and McDowell 1995. For arguments against the very coherence of the enterprise (which I think are unsuccessful), see Stich 1996, chapters 5 and 6.

[10] See *Euthyphro* 11a-b and surrounding passages, in Cooper (ed.) 1997.

[11] Crispin Wright gives a similar reading of the relevant bit of the *Euthyphro* in Wright 1992 (see 79-80). Interestingly, in a discussion later in the book (132fn), Wright takes himself to express partial disagreement with Socrates' principle. In fact, nothing he says there conflicts with it. His later discussion depends on a notion of "physical constitution" (132) such that physically constituting f-ness is not a matter of constituting f-ness (being what f-ness consists in). This is shown by his assertion that a dispositional-style view of color is consistent with taking redness to be "physically constituted" by non-dispositional physical features of a red object (131-132). What exactly physical constitution is supposed to be is left unclear.

[12] See also Anscombe 1979/1981.

[13] It is a complication I shall not pursue that what substitutes for "is an h" can even be a disjunctive predicate (i.e., of the form "is an l or an m or…"). If we do not want to allow disjunctive predicates in a statement of law, then we can weaken the requirement in the text as follows: all that is required for f's to reliably correlate with k's is that there be a set of predicates such that each member of the set generates a true statement when substituted for "is an h" in the schema, "It is a law that everything that is an h occurs in the presence of k's," and such that each

f satisfies at least one of the members of the set.

[14] What Fodor calls the "Representational Theory of the Mind" involves a commitment to the language-of-thought hypothesis (See Fodor 1990a). I am using a similar label to refer to a more minimal thesis, consistent with the language-of-thought hypothesis but also with other views about the physical carriers of mental content.

[15] For a discussion of a variety of suggestions of this sort, see Godfrey-Smith 1992. The terminology in scare quotes is from Field 1990 (see especially 106).

[16] This approach is most fully developed in Fodor 1990b. (But Fodor had already abandoned this approach by the time of the belated publication of the paper.)

[17] See Fodor 1990a. Fodor may have since backed away from this view as well; there is no mention of asymmetric dependency in his two later books on informational semantics, Fodor 1994 and Fodor 1998.

[18] Cowie, for example, writes, "I…want to emphasize that it is a *corollary* of the informational approach to concepts that possessing the concept of an F will usually involve one's knowing something about F's." (1990, 133, my emphasis). On the justification for Cowie's qualifying her claim with "usually", see the rest of this section.

[19] A comparable view is advanced by Aydede and Güzeldere (2005), who argue for the existence of "sensory concepts" the defining characteristic of which is that "their semantics is fixed…by a direct and immediate informational link to sensory experiences. To say that this link or sustaining mechanism is direct and immediate is to say that it is non-cognitive…" (p. 230). However, they caution that the array of concepts sustained non-cognitively is limited: "We have also shown that no concepts except sensory ones work this way—all others involve *cognitive* sustaining mechanisms—and that the rationale for this is a nomologically necessary fact about

autonomous intentional creatures like us. This is true even for what we have called perceptual concepts" (p. 231).

[20] There is a real question about whether this view can so much as be coherently formulated: we cannot just help ourselves to talk of intentional mechanisms in specifying a naturalistic account of content, and there are reasons for doubting that suitably naturalistic means exist for drawing the relevant boundary. But I will not pursue this issue here.

[21] Davidson 1963/1980, 14. The descriptions are logically related in the following sense: it is a logical truth that if there is a unique cause of B, the cause of B caused B.

[22] Many thanks to the following people for their helpful comments on versions of this paper: Marshall Abrams, Cheryl Chen, Jim Conant, Hannah Ginsborg, John Haugeland, Michael Kremer, John McFarlane, Elijah Milgram, and Barry Stroud.

References

Anscombe, G. E. M. (1979/1981) "'Under a Description'," *Metaphysics and the Philosophy of Mind*, Basil Blackwell, 208-219.

Aydede, Murat and Güven Güzeldere. (2005) "Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness," *Noûs* 39: 197-255.

Block, Ned. (1990) "Can the Mind Change the World?" *Meaning and Method: Essays in Honor of Hilary Putnam*, George Boolos (ed.), Cambridge University Press, 137-170.

Cowie, Fiona. (1999) *What's Within? Nativism Reconsidered*, Oxford University Press.

Cummins, Robert. (1989) *Meaning and Mental Representation*, MIT Press.

Davidson, Donald. (1963/1980) "Actions, Reasons and Causes," *Essays on Actions and Events*, Oxford University Press, 3-19.

Dretske, Fred. (1981/1999) *Knowledge and the Flow of Information*, CSLI Publications.

Dretske, Fred. (1990) "Replies to Reviewers," *Philosophy and Phenomenological Research* 50:

819-839.

Dretske, Fred. (1994) "If You Can't Make One, You Don't Know How It Works," *Midwest Studies in Philosophy*, v.19, Peter A. French, Theodore E. Uehling Jr. and Howard K. Wettstein (eds.), University of Minnesota Press, 468-482.

Field, Hartry. (1990) "'Narrow' Aspects of Intentionality and the Information-Theoretic Approach to Content," *Information, Semantics & Epistemology*, Enrique Villanueva (ed.), Blackwell, 102-116.

Fodor, Jerry. (1980) "Special Sciences, or the Disunity of Science as a Working Hypothesis," *Readings in the Philosophy of Psychology*, v.1, Ned Block (ed.), Harvard University Press, 120-133.

Fodor, Jerry. (1990a) *A Theory of Content and Other Essays*, MIT Press.

Fodor, Jerry. (1990b) "Psychosemantics: Or Where Do Truth Conditions Come From?" *Mind and Cognition*, William G. Lycan (ed.), Blackwell, 312-337.

Fodor, Jerry. (1994) *The Elm and the Expert*, MIT Press.

Fodor. Jerry, (1997) "Special Sciences: Still Autonomous After All These Years," *Philosophical Perspectives* 11: 149-163.

Fodor, Jerry. (1998) *Concepts: Where Cognitive Science Went Wrong*, Oxford University Press.

Fodor, Jerry. (1999) "Information and Representation," *Concepts: Core Readings*, Eric Margolis and Stephen Laurence (eds.), MIT Press, 513-524.

Glennan, Stuart. (2002) "Rethinking Mechanistic Explanation," *Philosophy of Science* 69: 342-353.

Godfrey-Smith, Peter. (1992) "Indication and Adaptation," *Synthese* 92: 283-312.

Kripke, Saul. (1980) *Naming and Necessity*, Harvard University Press.

Kripke, Saul. (1982) *Wittgenstein on Rules and Private Language*, Harvard University Press.

Margolis, Eric. (1999) "How to Acquire a Concept," *Concepts: Core Readings*, Eric Margolis and Stephen Laurence (eds.), MIT Press, 549-567.

McDowell, John. (1994) *Mind and World*, Harvard University Press.

McDowell, John. (1995) "Précis of *Mind and World*," *Philosophical Issues* 7: 231-239.

Millikan, Ruth Garrett. (1989/1991) "Biosemantics," *White Queen Psychology and Other Essays for Alice*, MIT Press, 83-101.

Plato. (1997) "Euthyphro," *Plato: Complete Works*, John Cooper (ed.). Hackett Publishing

Company, 1-16.

Stampe, Dennis. (1977) "Towards a Causal Theory of Linguistic Representation," *Midwest Studies in Philosophy*, v.2, Peter A. French, Theodore E. Uehling Jr. and Howard K. Wettstein (eds.), University of Minnesota Press, 42-63.

Stich, Stephen. (1996) *Deconstructing the Mind*, Oxford University Press.

Williamson, Timothy. (2000) *Knowledge and its Limits*, Oxford University Press.

Wright, Crispin. (1992) *Truth and Objectivity*, Harvard University Press.