Chapter 11
Causal/Informational Theories of Mental Content

Fred Adams

Abstract This entry looks at information-based causal theories of mental content. Causal theories appeal to causal conditions that exist between a representation (such as a thought or belief) and the part of the world represented (the content of the thought or belief). According to informational theories, mental states acquire their content by standing in appropriate informational and causal relations to objects (and properties) in the world. Very crudely, thoughts of dogs are about dogs (and mean *dog*) because information about dogs causes the thoughts that our minds use to keep track of dogs. This article explains some of the leading informational and causal theories of mental content—their twists, turns, refinements, and some of their leading criticisms (This article is a very focused history starting with Grice, Stampe, Dretske, and Fodor. There are many other important thinkers not discussed only for lack of space, not import. However, this is the history that started for me with my first graduate course at Wisconsin with Dennis Stampe just prior to his landmark paper. I later worked closely with Fred Dretske during the writing of his 1981 book and even later met and studied briefly with Jerry Fodor. I am grateful to Dretske, Fodor, and Stampe for their help over the years. I also thank Alfons Schuster for his patience, and an unidentified referee for quite useful advice.).

11.1 Introduction

This essay is about theories of mental content. What is the goal of such theories? It is to explain how minds can come into existence in purely physical systems and be explained via purely physical events and laws. At this point in the twenty-first century this may not be a surprising goal, but it has not always been so. For centuries (and even most of the first half of the last century) it was believed that there was something mysterious about minds that could not be captured by the physical sciences. This essay and the theories it evaluates, abandons that perspective. The

F. Adams (✉)
Department of Linguistics and Cognitive Science, University of Delaware Newark, 19716, Newark, DE, USA
e-mail: fa@udel.edu

expectations in the background here are that a mind is a purely physical mechanism and that meaning (one of those things earlier believed not to be the product of purely physical events) is the result of purely physical interactions between an organism and its environment. If one of these theories is true, we are purely physical and so are our minds. Understanding how we can think about the world with meaningful representations is the ultimate goal of the theories discussed below. We shall begin by asking how meaning comes into existence.

Once we can think, we can assign meaning, such as using the percent sign % as a symbol for percentages. But how are we able to think about percentages in order to assign symbols for them? How is the mind able to think about the world at all? Call the contents of our thoughts that are not assigned contents "unassigned meaning." What are the conditions of unassigned meaning, and how do purely physical systems acquire such meaning? The theories in this essay attempt to answer that question. They offer informational and causal conditions that a purely physical system could meet and in virtue of which acquire unassigned meaning. Thereby, thought and thought content is possible in purely physical, natural systems. In principle a computer would be able to think if it were to meet the causal conditions of the correct theory of mental content. These theories attempt to naturalize meaning—that is, not use the notion of meaning in the explanation of the acquisition of meaning, but use only natural causes and conditions.

I call these accounts "causal" and "informational." Of course, causation and information are not exactly the same notions, but they are related. They are not the same notions because there can be an informational relationship between events A and B, even when there is no direct causal relation between them (Dretske 1981). For example, when you and I are watching the same presidential debates on television in real time, I am receiving information on my television screen that tells me about what is on your television screen at the same time. However, there is no direct causal relation between your screen A and my screen B. There is a common cause C (the transmitters to each of our sets) that co-ordinates the events at A and B, even though there is not direct causal interaction between A and B.

What is more, causal interactions between events generate and convey information between the events. Fingerprints on the murder weapon carry information about who pulled the trigger. Rings in a tree carry information about seasons of growth. Rise and fall of barometric pressure carries information about high or low pressure fronts moving into the area.

In addition, these theories depend on a notion of information that is purely objective and part of the natural world. On this view, information is not mind-dependent. It is a feature of the natural world. Events happen against a background of what we may call a probability space. When it rains, there was a background probability of rain. When there is an earthquake it occurs against a background probability or likelihood of earthquakes in that area of the world. With the event's occurrence, there is a generation of information. There are various ways to measure that information (Shannon and Weaver 1949; Floridi 2016; Adams 2003b), but the import of information for the theories discussed below is the nomic relation between events and the world when information that a is F is carried by b's being G

(information that there is fire in the forest carried by smoke in the forest, information that a storm is pending by drop in barometric pressure). Dretske (1981, 1988) and Fodor (1984, 1987, 1990a,b,c) called it *information* or *indication*. It is this kind of informational relationship between natural events that is the basis of the accounts of mind and meaning discussed below, beginning with Grice's notion of natural information that he calls "natural meaning."

11.2   Natural vs. Non-natural Meaning

Paul Grice (1989) certainly deserves mention here. In "Meaning," (originally appearing in 1948 and again in 1957) Grice distinguished *natural meaning* from *non-natural meaning*. Natural meaning is of a sort that would not generate falsity. Where x naturally means that p, x "entails p" (Grice 1989, p. 213). If Colleen's spots naturally mean measles, then Colleen has measles. If smoke in the unspoiled forest naturally means fire then, given the presence of smoke, there is (or was) fire. The effect indicates or naturally means the cause.

   Grice gave the name *non-natural meaning* to things which have meaning but which can be false. Sentences can be false. I can say "Colleen currently has measles," where my utterance means but does "not entail" that Colleen currently has measles (Grice 1989, p. 214). Similarly, I can say "There is smoke in this forest" when there is no smoke or fire in this forest. Still, what I have said (though false) is perfectly meaningful. However, my utterances do not naturally mean or indicate measles or fire in the way that spots or smoke naturally mean or indicate them.

   Grice did not attempt to fully *naturalize* meaning. For he explained how the non-natural meaning of an utterance depends upon the content of the speaker's intentions and the audience's recognition of speaker intention. "'A meant (non-naturally) something by x' is (roughly) equivalent to 'A intended the utterance of x to produce some effect in an audience by means of the recognition of this intention (Grice 1989, p. 220).'" Grice did not further offer fully naturalized conditions of the origin of mental content of speaker's intentions or audience recognition, but had a major influence on those who did.

11.3   Isomorphism Plus Causation and Conditions of Fidelity

Dennis Stampe (1975, 1977, 1990) was one of the pioneers of causal/informational theories of content and representation. Influenced by Wittgenstein's (1961) picture theory of meaning, Stampe realized that to the Tractarian requirement of structural isomorphism between representation and represented one needs to add a causal requirement. Isomorphisms are symmetrical, but representations are not. "My thought was just that what was missing was causation. And then if you make

the thing represented not the state of affairs, the temperature is 70, but just the temperature, age of the tree, price of beans, etc., you can hold that the thing represented is the actual cause of the representation, and various determinate states of affairs causing various determinate representational states. And under ideal conditions or conditions of well-functioning, a certain identifiable state of affairs would cause a given representation, and that gives you the account of content and makes room for falsehood" [when things are not ideal$_{fa}$] (personal communication).

As for the classification of theories of content into causal accounts and teleological accounts, Stampe adds: "The idea that the specification of the relevant conditions ("fidelity" conditions, I called them) apparently requires bringing in reference to the function (teleology) of the representation-generating devices (not the representations themselves!) is put forward in Stampe (1977, 1990). So I take exception to some of the taxonomies of these theories that one sees" (personal correspondence).

For Stampe mental states (beliefs, desires, and intentions) both have representational objects (and contents) and are largely responsible for our success in the world. The harmony between representation and object represented cannot be accidental. "The idea that this determination is *causal* determination is, if not inevitable, only natural" (Stampe 1977, p. 82). Causal relations also explain *singularity* of representation. Consider two photographs of identical twins. What makes one a photo of *Judy* not of *Trudy* is not what is on the photographic film (there is no difference *in the photos*). It can only be that it was caused by *Judy* not *Trudy* that determines the singularity of representational content.

Still, what makes a representation of an object with property F *say that* the object is F? So the represented item and the representing item must co-exist and stand in some relation. "The hypothesis of my approach to the causal theory of content is that the theory of reference is a corollary of a causal theory of representation in general. This theory holds that it is by virtue of such a causal relation between representations generally and their objects that the latter are represented by the former. It may be that the thing represented is causally responsible for the representation [belief$_{fa}$] or vice versa [action$_{fa}$]..." (Stampe 1977, p. 84). All of this gives us the *representation of* relation, but we are stalking the *representation as* or *representation that* relation. How is that generated causally?

First, the appropriate causal relation for representation will hold between sets of properties F ($f_1 \ldots f_n$) of the thing represented (O), and a set of properties ($P_1 \ldots P_n$) of the thing doing the representing (R). The causal relation will establish an isomorphism between structures, between O's being F and R's being P. "The causal criterion requires that the relevant properties of the object represented cause the instantiation of the relevant properties in the...representation" (Stampe 1977, p. 85).

Second, if O's being F causes R to be P, then R is P only because O is F and wouldn't be P were it not for O's being F. That is, there is a nomic dependence of the sort that allows knowledge. Here, Stampe is influenced by Dretske's (1971) *conclusive reasons* theory of knowledge (where S knows that p on the basis of R when R wouldn't be the case unless p). When R wouldn't have P unless O was F,

R's being P tells one that O is F, on such an account of knowledge. So an example is, barring drought or other problems, one can tell from the rings in a tree how old it is.

Stampe calls the *represented as* properties the "expressed" properties of the representation. The expressed properties are the properties "the represented object would have if the representation of it were accurate," but he must "find a way to associate the concepts of accuracy and expression with natural processes and properties" (Stampe 1977, p. 87). Stampe appeals to "fidelity conditions." "If certain conditions do characterize those processes, the production of the representation would be caused by that state of affairs that the representation represents as being the case" (Stampe 1977, p. 89). The represented state does not have to be the actual cause of the representing state (or there would be no chance for falsity). In cases of natural representations (rings of a tree representing seasons of growth), fidelity conditions would include that there were no droughts, for instance. In cases of non-natural representations (linguistic utterances), fidelity conditions would include that the speaker does not intend to deceive, knows the language, and so on.

11.4   Information-Based Theories

Dretske's (1981) first attempt to explain mental content made use of the mathematical concept of information (Dretske 1981). For our purposes, it will suffice to say that a signal's carrying the *information* that *a* is *F*, like Grice's *natural meaning*, is a signal that cannot be false. If belief contents derive from information, as Dretske proposes, how can beliefs be false, when information cannot? This is one question Dretske set out to answer. Another was how a thought univocally could be that *a* is *F only*? For suppose that as a matter of scientific law, whatever is *F* is also *G*. Then no signal can carry the information that something is *F* without also carrying the information that something is *G*, just as nothing could naturally mean something is *F* without naturally meaning something is *G*. But Colleen's thought that Raven is a dog is only about being a *dog* (not about being a mammal, though Raven is both as a matter of natural law). Here we shall focus upon just these two matters.[1]

Take the second matter first. How can one develop a concept of dogs, when, upon receiving the information that something is a dog one is also receiving the information that something is a mammal? How can one's concept become specific to dogs alone (not mammals)? Dretske's answer is that it is due to what he calls *digitalization*. The basic idea is this. A signal that carries both the information that something is *F* and the information that something is *G*, carries the former piece of information in *digital* form if it carries the information that something is *G in virtue*

---

[1]For much more about information and its use in Dretske's philosophy, see Dretske (1983) and Adams (2003a,b, 2016).

of carrying the information that something is *F*. So if something is a dog, then it will be a mammal (and will be a mammal in virtue of being a dog, but not *vice versa*). Now one may be able to tell from the look and the bark that something is a dog, but not from that alone that it is a mammal (unless one already knows that dogs are mammals). So one may learn from the look and sound of dogs that they are dogs, but not (by look and sound alone) learn that the things seen and heard are mammals. Hence, one may acquire the concept *dog* without acquiring the concept *mammal*, even though one is getting both pieces of information.

Dretske goes on to say: "... we identify [a structure] *S*'s semantic content with its *outermost informational shell*, that piece of information *in which* all other information carried by *S* is nested.... This, of course, is merely another way of saying that *S*'s semantic content is... that piece of information *S* carries in digital form" (Dretske 1981, p. 178). Hence, a univocal concept is one that becomes selectively sensitive to a piece of information that something is *F* (carried in *digital form*). It is done when the content of a concept (of *Fs*) is caused by, informed by, a signal carrying a single piece of information (say that something is *F*) in completely digitalized form (Dretske 1981, p. 184).

We have reached a level where I can have a misrepresentation or false belief precisely because we have risen to a new level (a level beyond that of faithfully carrying information). This is Dretske's answer to Grice's challenge of going from *natural meaning* to *non-natural meaning* (at least for thought content, not utterances). Here is how Dretske puts it:

> But once we have meaning, once the subject has articulated a structure that is selectively sensitive to information about the F-ness of things, instances of this structure, tokens of the type, can be triggered by signals that *lack* the appropriate piece of information. When this occurs, the subject *believes* that *s* is *F* but, because this token of the structure type was not produced by the information that *s* is *F*, the subject falsely believes that *s* is *F*. We have a case of misrepresentation—at token of structure with false content. We have, in a word, meaning without truth (Dretske 1981, p. 195).

11.5   Attack on Wisconsin Semantics

Fodor's first foray into the world of naturalized semantics was to expose flaws in the accounts of Stampe and Dretske (both at Wisconsin at the time). He set *naturalistic* conditions on representation requiring, at a minimum, that "'*R represents S*' is *true iff C* where the vocabulary of *C* contains neither intentional nor semantic expressions" (Fodor 1990a, p. 32). While Fodor thinks the theories of Stampe and Dretske do not work, he emphasizes that "... something along the causal lines is the best hope we have for saving intentionalist theorizing, both in psychology and semantics" (Fodor 1990a, p. 34).

11.5.1 Contra Stampe

About *Wisconsin Semantics*, Fodor claims that there are "*two* Wisconsin theories about representation: one that's causal and one that's epistemic" (Fodor 1990a, p. 34). In support, Fodor gives this quote: "An object will represent or misrepresent the situation... only if it is such as to enable one to come to know the situation, i.e., what the situation is, should it be a faithful representation" (Stampe 1975, p. 223).

First, Fodor points out that what one can know from causal relations is non-symmetrical. Even though we can learn about the barometer (that it is low) from the weather (that it is storming), the weather doesn't represent the barometer. And second, Fodor approvingly agrees with Stampe (1977) that the epistemic account founders on the "singularity" of representation. Stampe at one point discusses a xerox machine making multiple copies. From each copy one can learn something about the other copies, but the copies do not represent one another. They only represent the one original—so only causal conditions will tease out the right representational object, according to Stampe.

Fodor believes Stampe's theory mishandles misrepresentation (because it is epistemic). Fodor gives this quote from Stampe (1975, p. 223):

> An object will represent or misrepresent the situation... only if it is such as to enable one to come to know the situation, i.e., what the situation is, should it be a faithful representation. If it is not faithful, it will misrepresent the situation. That is, one *may* not be able to tell from it what the situation is, despite the fact that it is a representation of the situation. In either case, it represents the same thing, just as a faithful and unrecognizable portrait may portray the same person.

Fodor maintains that this gets things the wrong way around. It is not that something is a faithful portrait, say of Mao, because one can learn something about Mao from it. Rather, one can learn something from it about Mao because it is a (faithful) portrait of Mao. Fodor also maintains that there is a "nasty scope ambiguity" (Fodor 1990a, p. 36) between:

(a) if R is faithful (you can tell what the case is); vs.
(b) you can tell (what the case is if R is faithful).

Despite this, Fodor admits it is clear "that it is (a) that Stampe intends...." (Fodor 1990a, p. 36), and turns to the following example. Suppose that Tom is Swiss. Then suppose Denny says: "Tom is Armenian." Fodor says that Stampe maintains that the sentence represents (i.e, misrepresents) Tom's being Swiss because that is the fact to which, if faithful, the representation would provide epistemic access. Fodor maintains that there is no clear way to understand this claim. He claims that the only ways the sentence could be faithful would be to change the facts (make Tom Armenian) or change the sentence (say "Tom is Swiss"). There is further evidence that this is the general tenor of Fodor's complaint when he later discusses another of Stampe's examples (with much the same upshot over the disjunction problem). Consider the following longish quote from Stampe (1977, p. 49):

> The number of rings (in a tree stump) represents the age of the tree.... The causal
> conditions, determining the production of this representation, are most saliently the climatic
> conditions that prevailed during the growth of the tree. If these are normal...then one ring
> will be added each year. Now what *is* that reading.... It is not, for one thing, infallible. There
> may have been drought years.... It is a *conditional* hypothesis: that *if* certain conditions
> hold, then something's having such and such properties would cause the representation
> to have such and such properties.... Even under those normal conditions, there may be
> other things that would produce the rings—an army of some kind of borer, maybe, or an
> omnipotent evil tree demon.

As Fodor rightly points out, Stampe has to make a decision about what is "in"
and what is "out" of normalcy. Why, for instance, are droughts abnormal but not
borers? "And, of course, given Stampe's decision, it's going to follow from the
theory that the tree's rings represent the tree's age and that the tree-borer-caused
tree rings tokens are wild (i.e., that they *mis*represent the tree's age). The worrying
question is what, if anything, motivates this decision" (Fodor 1990a, p. 44–45).
Fodor accuses Stampe of not having a principled way of deciding the matter. Further,
he claims that Stampe does not give a principled reason for deciding what counts as
a *ring*. The borer marks look like rings, but that why do they count *as rings*? But
worst of all, Fodor's claims indicate that Stampe has not moved beyond natural
meaning. For there is no principled way to say when a "ring" is wild. Do the
"rings" produced by borers *mis*represent seasons, or veridically represent *seasons
or borers*? Without principled answers to the former, there is no principled answer
to the latter questions, and no solution to the disjunction problem (immediately
below).

Now Stampe does appeal to teleology and function in his account of representa-
tion. And Fodor admits (at least here, though he would deny it in other writings) that
if one could show there was a teleological mechanism that only produced genuine
rings in growth seasons, and wild tokens of *rings* when "Mother Nature is a little
tipsy", and things that look like but are not rings, when borers are at work, then we
could have misrepresentation (for the wild tokens). But this puts weight on teleology
that it will not bear.

11.5.2 Contra Dretske

Fodor's main complaint against Dretske's theory of 1981 is Dretske's use of the
learning period L to try to solve the "disjunction problem" (and explain how
misrepresentation occurs). He reminds us that Dretske's "...way out of the problem
about disjunction is to enforce a strict distinction between what happens in the
learning period and what happens after" (Fodor 1990a, p. 40). If we call a "wild"
tokening of a concept one caused by something that is not represented by the
concept, then wild tokenings are those that are uncorrected by the teacher and that
happen *after* the learning period...and thus, after concept formation. Hence, wild
tokenings are misrepresentations, on Dretske's view.

Fodor's reaction is classic. He says: "This move is ingenious but hopeless. . . . Just for starters, the distinction between what happens in the learning period and what happens thereafter surely isn't principled: there is no time after which one's use of a symbol stops being merely shaped and starts to be, as it were, in earnest" (Fodor 1990a, p. 41).

Even if we could draw a determinate line between what is inside and outside the learning period (a time line), Fodor claims that the account still "doesn't work" because "it ignores relevant counterfactuals" (Fodor 1990a, p. 41). Let's consider the fox example to explain. Dretske would say my concept "fox" has foxes as its content because during the learning period it was trained on foxes. The teacher conditioned my "fox" symbol to fire in the presence of foxes (on the information that something is a fox) and to extinguish on non-foxes (on the information that something is a non-fox). So my "fox" symbol became perfectly correlated with foxes in the learning period *L*. But Fodor asks what *would have* happened *in L*, had a sheltie with a trim been shown to me? Since we know by stipulation that *after L*, a sheltie with a trim would fire my "fox" symbol, we know shelties (the information that something is a sheltie) would activate my concept "fox." So we know that the information that something is fox would fire my "fox" concept. And we know that the information that something is a sheltie (with a trim) would fire my "fox" concept. So exactly what information was I sensitive to *during the learning period L*? Was it the information *that something is a fox*? Or was it the information *that something is a fox or sheltie*? Fodor claims that the appeal to the learning period *L* alone does not answer this question. The firing of my "fox" symbol is as well explained by either piece of information (indeed better explained by the latter), and if it is the latter disjunctive piece, then when I see a sheltie and say "fox" I do not misrepresent. For my thought content would be true. "Fox" for me would mean *fox or sheltie*, despite all diligence taken by the teacher during Dretske's learning period *L*. Thus, Fodor claims that Dretske has not solved the *disjunction problem* after all.

Later in this particular article (Fodor 1984), Fodor considers appealing to tokenings of symbols under *normal circumstances* (p. 42) as a way out, but later (Fodor 1990a,b) comes to reject this approach too as hopeless. Fodor also flirts briefly with the idea of tying an account of meaning to *teleology* (Fodor 1990a, p. 43; 1990c), but drops this idea entirely later (Fodor 1990a,b). The idea was that *R's* represents *S's*, when *S's* cause *R's* under normal conditions (or when *S's* are supposed to cause *R's*, i.e., when it is their function to do so). It is not clear whether he was considering these approaches on his own or just in thinking of the theories of Stampe and Dretske. In any case, he had thoroughly discarded these approaches in a very short time (Fodor 1987, 1990a,b), in part because "representations generated in teleologically normal circumstances must be true" (Fodor 1990a, p. 47). So he didn't foresee these considerations giving a happy account of misrepresentation.

11.6   Dretske's Response: Indicator Function Account

In 1988 Dretske revised his account of naturalized meaning.[2] He changed and
simplified the account. Dretske's new recipe for content involves three interlocking
pieces. (1) The content of a symbol "C" must be tied to its natural meaning
F (Fs—objects that are F). (2) Natural meaning (indication, information) must
be transformed to semantic content. The transformation of acquired information
content into cognitive (semantic) content—must be encoded in a form capable of
being harnessed to beliefs and desires in service of the production of behavior
M. (3) The causal explanation of the resultant behavior M must be in virtue of
the informational content of the input states. Thus, if a symbol "C" causes bodily
movements M because tokenings of "C" indicate (naturally mean) Fs, then "C"
is elevated from merely naturally meaning Fs to having the semantic content that
something is F.

$$F \leftarrow \text{ indicates "C" and causes } \rightarrow M \text{ (because it indicates F)}$$

Dretske's account is essentially historical. In different environments, the same
physical natural signs may signify different things, and have different natural
meaning. On Earth, Al's fingerprints are natural signs or indicators of Al's presence.
On Twin-Earth, the same physical types of prints indicate Twin-Al's presence, not
Al's. For this to be true, there must be something like an *ecological boundary*[3] that
screens off what is possible in one environment from what is possible in another.

Dretske's solution to the disjunction problem has at least two components. The
symbol "C" must start out with the ability to naturally mean Fs. Even if "C"s
indicate Fs only, to acquire semantic content, a symbol must lose its guarantee of
possessing natural meaning. It needs to become locked to Fs and permit robust, and
even false, tokening, without infecting its semantic content. A "learning period"
doesn't quite work, unembellished. So Dretske now appeals to the explanatory
relevance of the natural meaning. For Dretske, it is not just what causes "C"s, but
what "C"s in turn cause, and why they cause this that is important in locking "C"s
to their content (F).

Let's suppose that a ground squirrel needs to detect Fs (predators) to stay alive.
If Fs cause "C"s in the ground squirrel, then the tokening of "C"s indicate Fs.
Dretske claims that "C"s come to have the content that something is an F, when
"C"s come to have the function of indicating the presence of Fs. When will that
be? For every predator is not just a predator, it is an animal (G), a physical object
(H), a living being (I), and so on for many properties. Hence, tokens of "C" will
indicate all of these, not just Fs. Dretske's answer is that when "C"s indication of Fs

---

[2]Many people (mistakenly) still think that Dretske's considered answer to the "disjunction
problem" and misrepresentation is his account in his 1986 paper (Dretske 1986). That is incorrect.
He abandoned that account in favor of the solution in his 1988 book (Dretske 1988, chapter 4).

[3]Let's think of an ecological boundary as akin to what Dretske (1981) calls a "channel condition."

(alone) *explains* the animal's behavior, then "C"s acquire the semantic content that something is a predator (F). Hence, it is the *intensionality* of explanatory role that locks "C"s to F, not to G or H or I.

For Dretske, behavior is a complex of a mental state's causing a bodily movement. So when "C" causes some bodily movement M (say, the animal's movement into its hole), the animal's movement consists of its trajectory into its hole. The animal's behavior is its causing that trajectory. The animal's behavior—running into its hole—consists of "C"s causing M ("C" → M). There is no specific behavior that is required to acquire an indicator function. Sometimes the animal slips into its hole (M1). Sometimes it freezes (M2). Sometimes it scurries away (M3). This account says that "C"s become *recruited* to cause such movements because of what "C"s indicate (naturally mean). The animal needs to keep track of Fs and it needs to behave appropriately in the presence of Fs (to avoid predation). Hence, the animal thinks there is a predator when its token "C" causes some appropriate movement M (and hence the animal behaves) because of "C"s indication (natural meaning). Not until "C"s natural meaning *has an explanatory role* does "C" lock to its semantic content F. So "C"s acquired function to indicate or detect predators elevates its content to the next, semantic level. Now "C" can be falsely or otherwise robustly tokened. The animal may run into its hole because it thinks there is a predator, even when spooked only by a sound or a shadow, as long as the presence of sounds or shadows doesn't explain why the "C"s cause relevant Ms (don't explain the animal's behavior).[4]

On this view, indicator functions are like other natural functions, such as the function of the heart or kidneys or perceptual mechanisms. The account of natural functions favored by Dretske is one on which the X acquires a function to do Y when doing Y contributes some positive effect or benefit to an organism and so doing helps explain why the organism survives. Then there is a type of selection for organisms with Xs that do Y. Consequently, part of the reason X's are still present, still doing Y, is that a type of selection for such organisms has taken place. Of course, this doesn't explain how X got there or began doing Y, in the beginning.

Naturally, the selection for indicator functions has to be within an organism's lifetime, not across generations. Dretske thinks of this kind of selection as a type of biological process of "recruitment" or "learning" that conforms with standard, etiological models of natural functions (Adams 1979; Adams and Enc 1988; Enc and Adams 1998).

Now the last piece of the puzzle is to show that the content of "C" at some level is relevant to the explanation of the organism's behavior. "C" may cause M, but not because of its natural meaning. "C"s meaning may be idle. For this purpose, Dretske distinguishes *triggering* and *structuring* causes. A triggering cause may be the thing that causes "C" to cause M right now. Whereas a structuring cause is what

---

[4]Sticks and stones may break one's bones, but shadows and sounds cannot harm you. Every ground squirrel knows this. So no "C" is recruited to be an indicator of shadows or sounds. Predators. . . that is altogether different.

explains why "C" causes M, rather than some other movement N. Or, alternatively, structuring causes may explain why it is "C" rather than some other state of the brain D that causes M. So structuring causes highlight *contrastives*: (a) why "Cs" cause M, or alternatively, (b) why "C"s cause M. In either case if it is because of "C"s natural meaning, then we have a case of structuring causation, and content plays a role on this account of meaning mechanisms.

11.7 Fodor's Asymmetrical Causal Dependency Theory of Meaning

Fodor (1987, 1990a,b, 1994) offers conditions sufficient for a symbol "X" to mean something X.[5] Let's also be clear that Fodor is offering conditions for the meanings of primitive, non-logical thought symbols. This may well be part of the explanation of why he sees his conditions as only sufficient for meaning. The logical symbols and some other thought symbols may come by their meanings differently. Symbols with non-primitive (molecular) content may derive from primitive or atomic symbols by decomposing into atomic clusters. It is an empirical question when something is a primitive term, and Fodor is the first to recognize this.

Fodor's conditions have changed over time and are not listed by him anywhere in the exact form below, but I believe this to be the best representation of his current[6] considered theory. (This version is culled from Fodor (1987, 1990a,b, 1994).) The theory says that "X" means X if:

(1) 'Xs cause "X"s' is a law,
(2) For all Ys not = Xs, if Ys qua Ys actually cause "X"s, then Y's causing "X"s is asymmetrically dependent on Xs causing "X"s,
(3) There are some non–X–caused "X"s,
(4) The dependence in (2) is *synchronic* (not diachronic).

Condition (1) represents Fodor's version of natural meaning (information, indication). If it is a law that Xs cause "X"s, then a tokened "X" may indicate an X. Whether it does will depend on one's environment and its laws, but this condition affords[7] natural meaning a role to play in this meaning mechanism. For "X" to become a symbol for Xs requires more than being tokened by Xs. "X"s must be dedicated to, faithful to, locked to Xs *for their content*.

Condition (2) is designed to capture the jump from natural meaning to semantic content and solve the disjunction problem. It does the work of Dretske's learning

---

[5]Fodor's conditions for meaning are in flux and (subtly) change across these three works.

[6]Below we will consider another incarnation of the theory that adds a condition and discuss why he may have added and then dropped that condition. For more about this see Adams and Aizawa (1994).

[7]Fodor likes to refer to his view as an "informational" semantics (Fodor 1994).

period, giving us a new mechanism for locking "X"s to Xs. Fodor's fix is to make all non-X-tokenings of "X"s nomically dependent upon X-tokenings of "X"s from the very start. There is then no need[8] for a learning period. The condition says that not only will there be a law connecting a symbol "X" with its content X, but for any other items that are lawfully connected with the symbol "X", there is an asymmetrical dependency of laws or connections. The asymmetry is such that, while other things (Ys) are capable of causing the symbol to be tokened, the Y → "X" law depends upon the X → "X" law, but not *vice versa*. Hence, the asymmetrical dependence *of laws* locks the symbol to its content.

Condition (3) establishes "robust" tokening (Fodor 1990b). It acknowledges that there are non-X-caused "X"s. Some of these are due to false thought content, as when I mistake a horse on a dark night for a cow, and falsely token "cow" (believing that there is a cow present). Others are due to mere associations, as when one associates things found on a farm with cows and tokens "cow," (but not a case of false belief). These tokenings do not corrupt the meaning of "cow" because "cow" is dedicated to cows in virtue of condition (2).

Condition (4) is designed to circumvent potential problems due to kinds of asymmetrical dependence that are not meaning conferring (Fodor 1987, p. 109). Consider Pavlovian conditioning. Food causes salivation in the dog. Then a bell causes salivation in the dog. It is likely that the bell causes salivation only because the food causes it. Yet, salivation hardly means food. It may well naturally mean that food is present, but it is not a thought or thought content and it is not ripe for false semantic tokening. Condition (4) allows Fodor to block saying that salivation[9] itself has the semantic content that food is present, for its bell-caused dependency upon its food-caused dependency is diachronic, not synchronic. First there is the unconditioned response to the unconditioned stimulus, then, over time, there comes to be the conditioned response to the conditioning stimulus. Fodor's stipulation that the dependencies be synchronic not diachronic screens off Pavlovian conditioning and many other types of diachronic dependencies, as well.

11.8 Conclusion

There are many problems with Fodor's theory that I have detailed elsewhere and for which there is not space to include here (Adams 2003a; Adams and Aizawa 2010). I cannot help but believe that causal/informational theories of mental content have to be correct, in the end. If our ability to cognitively interact with our environments is not magic, or ultimately inexplicable, then the explanation of our cognitive abilities

---

[8]Nor is there a need for learning (period)—consistent with Fodor's penchant for nativism.

[9]One might think that it doesn't need blocking because salivation is not a vehicle in the language of thought (LOT). But Fodor does not restrict his theory to items in LOT. So in principle, even things outside the head can have meaning.

must rest with our causal and informational interactions with our environment. While it is true that there are alternative theories of mental content (causal role theories and teleological theories, for example), these theories too exploit causal and informational relations between organism and environment to explain the origin of mental content. The difference in these theories and the ones we have discussed ultimately are differences in the kinds of causal explanations required for mental content—not that some accept and some reject causal conditions as necessary for mental content. Further, as we have seen, some (such as Stampe) would even reject the taxonomic division of causal theories of mental content versus teleological theories, from the start (in sharp contrast with Fodor, of course). As discussion moves forward on the nature of the mind and cognition in philosophy of mind and cognitive science, generally, I predict that progress will be in the form of answering the worries and objections of the last section. I do not believe we will see progress made by moving away from causal and informational theories of mental content. That is why in Adams (2003b), I said there is no going back.

References

Adams F (1979) A goal-state theory of function attribution. Can J Philos 9(3):493–518

Adams F (2003a) Thoughts and their contents: naturalized semantics. In: Stich SP, Warfield TA (eds) The Blackwell guide to philosophy of mind. Basil Blackwell, Oxford, pp 143–171

Adams F (2003b) The informational turn in philosophy. Mind Mach 13(4):471–501

Adams F (2016) Information and cognition. In: Floridi L (ed) The Routledge handbook of philosophy of information. Routledge, New York, pp 345–356

Adams F, Aizawa K (1994) "X" means X: Fodor/Warfield semantics. Mind Mach 4(2):215–231

Adams F, Aizawa K (2010) Causal theories of mental content. Stanf Encycl Philos. Available via http://plato.stanford.edu/entries/content-causal/

Adams F, Enc B (1988) Not quite by accident. Dialogue 27(2):287–297

Dretske F (1971) Conclusive reasons. Australas J Philos 49(1):1–22

Dretske F (1981) Knowledge and the flow of information. MIT/Bradford Press, Cambridge

Dretske F (1983) Precis of knowledge and the flow of information. Behav Brain Sci 6(1):55–63

Dretske F (1986) Misrepresentation. In: Bogdan RJ (ed) Belief. Oxford University Press, Oxford, pp 17–36

Dretske F (1988) Explaining behavior: reasons in a world of causes. MIT/Bradford, Cambridge

Enc B, Adams F (1998) Functions and goal-directedness. In: Allen C, Bekoff M, Lauder G (eds) Nature's purposes. MIT/Bradford, Cambridge

Fodor JA (1984) Semantics, Wisconsin style. Synthese 59(3):231–250. (Reprinted in Fodor, 1990)

Fodor JA (1987) Psychosemantics: the problem of meaning in the philosophy of mind. MIT/Bradford, Cambridge

Fodor JA (1990a) A theory of content and other essays. MIT/Bradford Press, Cambridge

Fodor JA (1990b) Information and representation. In: Hanson PP (ed) Information, language, and cognition. University of British Columbia Press, Vancouver

Fodor JA (1990c) Psychosemantics or: where do truth conditions come from? In: Lycan WG (ed) Mind and cognition. Basil Blackwell, Oxford

Fodor JA (1994) The Elm and the expert. MIT/Bradford, Cambridge

Floridi L (2016) The Routledge handbook of philosophy of information. Routledge, New York/London

Grice P (1989) Studies in the way of words. Harvard University Press, Cambridge

Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana/Chicago

Stampe DW (1975) Show and tell. In: Freed B, Marras A, Maynard P (eds) Forms of representation. North-Holland, Amsterdam, pp 221–245

Stampe DW (1977) Toward a causal theory of linguistic representation. Midwest Stud Philos 2(1):42–63

Stampe DW (1990) Content, context, and explanation. In: Villanueva E (ed) Information, semantics, and epistemology. Basil Blackwell, Oxford, pp 134–152

Wittgenstein L (1961) Tractatus Logico-Philosophicus. Routledge & Kegan Paul, London

Other Relevant Resources

(1) Adams F, Aizawa K (1992) "X" means X: semantics Fodor-style. Mind Mach 2(2):175–183

(2) Adams F, Aizawa K (1994) Fodorian semantics. In: Stich SP, Warfield TA (eds) Mental representation. Basil Blackwell, Oxford, pp 223–242

(3) Adams F, Drebushenko D, Fuller G, Stecker R (1990) Narrow content: Fodor's folly. Mind Lang 5(3):213–229

(4) Adams F, Dietrich LA (2004) What's in a(n empty) name? Pac Philos Q 85(2):125–148

(5) Adams F, Stecker R (1994) Vacuous singular terms. Mind Lang 9(4):387–401

(6) Antony LM, Levine J (1991) The nomic and the robust. In: Loewer BM, Rey G (eds) Meaning in mind: Fodor and his critics. Basil Blackwell, Oxford, pp 1–16

(7) Baker LR (1989) On a causal theory of content. Philos Perspect 3:165–186

(8) Baker LR (1991) Has content been naturalized? In: Loewer BM, Rey G (eds) Meaning in mind: Fodor and his critics. Basil Blackwell, Oxford, pp 17–32

(9) Bar-On D (1995) "Meaning" reconstructed: Grice and the naturalizing of semantics. Pac Philos Q 76(2):83–116

(10) Boghossian PA (1991) Naturalizing content. In: Loewer BM, Rey G (eds) Meaning in mind: Fodor and his critics. Basil Blackwell, Oxford, pp 65–86

(11) Cummins R (1989) Meaning and mental representation. MIT/Bradford, Cambridge

(12) Dennett DC (1987) Review of J. Fodor's psychosemantics. J Philos 85:384–389

(13) Enc B (1982) Intentional states of mechanical devices. Mind 91:161–182

(14) Enc B (2002) Indeterminacy of function attributions. In: Ariew A, Cummins R, Perlman M (eds) Functions: new essays in the philosophy of psychology and biology. Oxford University Press, pp 291–313

(15) Fodor JA (1991) Replies. In: Loewer BM, Rey G (eds) Meaning in mind: Fodor and his critics. Basil Blackwell, Oxford, pp 255–319

(16) Fodor JA (1998) Concepts: where cognitive science went wrong. Oxford University Press, Oxford

(17) Fodor JA (1998) In critical condition: polemical essays on cognitive science and the philosophy of mind. MIT/Bradford Press, Cambridge

(18) Godfrey-Smith P (1989) Misinformation. Can J Philos 19(4):533–550

(19) Godfrey-Smith P (1992) Indication and adaptation. Synthese 92(2):283–312

(20) Jones T, Mulaire E, Stich S (1991) Staving off catastrophe: a critical notice of Jerry Fodor's psychosemantics. Mind Lang 6(1):58–82

(21) Loar B (1991) Can we explain intentionality? In: Loewer BM, Rey G (eds) Meaning in mind: Fodor and his critics. Basil Blackwell, Oxford, pp 119–135

(22) Maloney CJ (1990) Mental misrepresentation. Philos Sci 57(3):445–458

(23) Manfredi PA, Summerfield DM (1992) Robustness without asymmetry: a flaw in Fodor's theory of content. Philos Stud 66(3):261–283

(24) Possin K (1988) Sticky problems with Stampe on representations. Australas J Philos 66(1):75–82

(25) Stampe DW (1986) Verification and a causal account of meaning. Synthese 69(1):107–137
(26) Sterelny K (1990) The representational theory of mind. John Wiley & Sons
(27) Wallis C (1994) Representation and the imperfect ideal. Philos Sci 61(3):407–428
(28) Wallis C (1995) Asymmetrical dependence, representation, and cognitive science. South
      J Philos 33(3):373–401
(29) Warfield TA (1994) Fodorian semantics: a reply to Adams and Aizawa. Mind Mach 4(2):
      205–214
(30) Wright L (1973) Functions. Philos Rev 82(2):139–168

- Field Guide to the Philosophy of Mind Entry on Fodor's Asymmetrical Causal Dependency
  Theory of Meaning (and related entries on thought and language). Available via
  http://host.uniroma3.it/progetti/kant/field/asd.htm.
- Teleological Theories of Mental Content (Ruth Millikan). Available via
  http://www.ucc.uconn.edu/~wwwphil/Teleocnt.pdf.
- Externalism about mental content, Intentionality, Language of thought hypothesis, Narrow
  mental content, Non-conceptual mental content, Representational theories of consciousness,
  Teleological theories of mental content.