

Fodor and Psychological Explanations *

John Perry[†] and David Israel[‡]

[In *Meaning in Mind*, edited by Barry Loewer and Georges Rey. Oxford: Basil Blackwell, 1991, 165–180. Reprinted in John Perry, *The Problem of the Essential Indexical and Other Essays*. New York: Oxford University Press, 1993.]

1 Introduction

1.1 The Texts

We begin with some quotations from Fodor.

[I]t is crucial to the whole program of explaining behavior by reference to mental states that the propositional attitudes belonging to these [causal] chains are typically non-arbitrarily related in respect to their content (taking “content” of a propositional attitude, informally, to be whatever it is that the complement of the corresponding PA-ascribing sentence expresses). One can imagine the occurrence of causal chains of mental states which are not otherwise related (as, e.g., a thought that two is a prime number, causing a desire for tea...)...Still if all our mental life were like this, it’s hard to see what point ascriptions of contents to mental states would have...The paradigm situation—the grist for the cognitivist’s mill—is the one where the propositional attitudes interact causally and do so in virtue of their content. And the paradigm of this paradigm is the practical syllogism...John believes that it will rain if he washes his car. John wants it to rain. So John acts in a manner intended to be a car-washing...Our common-sense psychological generalizations relate mental states in virtue of their content...(*Propositional Attitudes*, as published in **Representations**, pp.182-184)

I dearly wish that I could leave this topic here, because it would be very convenient to be able to say, without qualification, what I strongly implied

*The research reported in this paper has been made possible by a gift from the System Development Foundation.

[†]Philosophy Department, Stanford University and Center for the Study of Language and Information(CSLI)

[‡]Artificial Intelligence Center, SRI International and CSLI

above: the opaque readings of propositional attitude ascriptions tell us how people represent the objects of their propositional attitudes...(*Methodological Solipsism, Representations*, p. 236)

What I think is exactly right is that the construal of propositional attitudes which such a psychology renders is nontransparent...The trouble is that nontransparency isn't quite the same notion as opacity, as we shall now see. (*ibid*, p. 236)

Having said all this, I now propose largely to ignore it...(*ibid*, p. 239)

2 The Contents and Causal Powers of Tokens of Mentalese

Fodor takes propositional attitudes to be relations to tokens of an internal language (“mentalese”) that have content. If Jerry believes that *P*, Jerry has a token of mentalese in his belief structure that has the content that *P*. These tokens have causal properties as well as content; they are related in law-like ways in virtue of their causal properties. But the laws must (in the paradigm cases that are grist for the cognitivist’s mill) make sense in terms of the contents of the tokens. It is the contents that are related in commonsense psychological principles, and it is the meshing of content and causal properties that makes it conceivable that cognitive psychology might work.

What exactly is the relation between the contents and causal properties of tokens? The first quotation above, from *Propositional Attitudes*, suggests a simple answer: tokens interact as they do in virtue of their content. But Fodor begins *Methodological Solipsism* with a quotation from Hume that states a problem for this view.

...to form the idea of an object and to form an idea simply is the same thing; the reference of the idea to an object being an extraneous denomination, of which in itself it bears no mark or character.¹

The fact that my idea of red is an idea of one color rather than another is an *external denomination*—a relation between the idea and a color—not something that can influence the way it interacts with other ideas. The view that emerges in *Methodological Solipsism* is not that tokens interact in virtue of their contents, but that both the causal facts and the content facts about a token are settled by its *formal* properties. The formal properties are, roughly, those that a processor scanning the tokens can detect and by which its actions can be systematically determined. If both the contents and the causal powers of a token depend on its form, the two kinds of properties might mesh in the ways necessary to have the content-based principles of cognitive psychology (or commonsense psychology) backed by causal laws relating formally individuated

¹*Representations*, p.225, quoting from Book I of Hume’s *Treatise*.

states. One imagines a species-wide causal role for a given type of token and a species-wide interpretive function that assigns contents to types. The two mesh so that, for example, within the human species, if a belief and a desire cause a volition, then the contents of the belief, desire, and volition have some sensible, rational connection with one another.

One imagines the interpretive function fixing the reference of a certain formal type as the property of being red, once and for all—the same for each member of a cognizing species. But it seems that the reference of many of our ideas is not only external, but *circumstantial*. The interpretive function and the formal properties of a token do not fully determine the reference; particular circumstances that vary among formally identical tokens, also must be taken into account. Suppose, for example, that Jerry is looking at a cup. The cup that he is looking at will be the referent of the mentalese phrases that are analogues to “the cup in front of me” or “that cup.” Tokens of the same mentalese phrases, in the head of someone else, looking at a different cup, will have different referents.²

How can causal and content properties of tokens mesh, if the content properties of a token depend both on its form and on such particular external circumstances, while the causal properties depend on form alone? If contents are sensitive to external circumstances, and so classify persons who are internally similar as different, and those who are internally different as similar, how can content-based principles of rationality mesh with causal laws?

If such circumstantially determined external references of phrases become part of the truth-conditions of the sentences in which they occur, then those truth conditions will also be infected with circumstantiality. It seems a causally coherent psychology should not individuate states in terms of such external denominations. But the contents that are ascribed to mental states by our practices of propositional attitude reporting are often based on just such circumstantially determined external denominations. The contents that a person believes and the people who are counted as believing the same thing depend in part on external facts about reference and hence on nonformal properties of their tokens. As a result, people with formally identical tokens may believe different things; people with formally differing tokens may believe the same thing.

Fodor would have liked the following response to this worry to work:

Transparent reports of attitudes do classify mental states in a circumstantial, nonformal way. But because this is so, we do not take transparent reports to tell us how the agent is thinking, and for this reason explanations using transparent reports are impotent. But opaque reports are explanatorily valuable, because they do tell us how agents think; they can do this because they don't classify mental states circumstantially, but in terms of contents that mesh with formal properties.

But in the remarks quoted above from *Methodological Solipsism*, he indicates that

²See below §3, where we argue for such analogues, and §4.2, where we sketch a semantics for mentalese in which the relation between English sentences and sentences of mentalese is taken to be identity.

this response is not quite right. Even opaque reports rely on circumstantial classification. In effect, he postulates a level of *fully opaque (nontransparent)* content that works the way he had hoped the opaque classification would. But he does not provide very many details.³

In this paper we sketch a theory, within a Fodorian conception of cognition, of how the formal properties and external circumstances of tokens relate to the contents of tokens, and of how commonsense explanations might work given this framework. We conclude that

- Circumstantially determined content properties can mesh with causal properties.
- Attitude reports in terms of circumstantially determined content can be explanatory, that is, can enter into adequate psychological explanations.
- It is scarcely conceivable that some content properties not be circumstantially determined.

Like Fodor, we believe cognitive psychology is possible, and that there are strong empirical arguments in favor, and only weak and avoidable arguments against, the thesis that a *component* of that psychology will be individualistic. We think of ourselves as providing a way of looking at content and content-based explanations that should be helpful to those, like Fodor and ourselves, who believe cognitive psychology can and should include such an individualistic component. There are nevertheless some points of disagreement between us and Fodor on the topics covered in this essay. We think transparent explanations can work well; Fodor does not. In our reconstruction, even fully opaque attitude reports will be somewhat circumstantial. We provide a completely noncircumstantial level of content, but claim that to understand the rationality of laws one needs to bring in the (admittedly rather modest) circumstance that the relevant tokens of mentalese belong to the same agent.

2.1 A Broader Context of Disagreement

Given the amicable goal of this essay, perhaps we should indicate some broad areas of disagreement before plunging in. Like Fodor, we conceive of individualistic cognitive psychology as requiring structured internal states that have both content and causal role. And like him, we believe that the postulation of such states is empirically reasonable and that the success and structure of our ordinary psychological concepts provides important evidence for it. We are inclined, however, to think that Fodor's talk of the language of thought and mentalese may encourage the assumption that the internal

³The present essay is the result of years of arguing about essays in **Representations**, particularly *Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology*. Many of the most fruitful of these arguments have involved Lisa Hall and Brian Smith, whom we thank. Special thanks to Georges Rey for a thoughtful and illuminating critique of an earlier draft. We intend the present piece as an outward and visible symbol of the inner cycles of inspiration and frustration we owe to that essay. We have not been able to pay as much attention to Fodor's later work **Psychosemantics** as would have been optimal.

structures are more language-like than there is any reason to believe. And we worry about whether the notion of *formality* coalesces conditions from a number of different areas that have less in common than Fodor thinks that they do.⁴ But for the purposes of this paper we set aside these worries.

Other areas of disagreement concern questions of how the causal and content properties of internal states relate to one another. An important issue is the source of what we shall call the interpretive function, which assigns contents to formally individuated mental states. We sympathize with the aim and intent of approaches that see content properties as ultimately reducible (in some weak sense) to causal properties, with a role for information “Wisconsin style.” But we do not think that any such reduction to causal/informational concepts can succeed. Our own view is a sort of naturalistic functionalism that stresses *functions* and *purposes*—hence the results of actions measured in terms of degree of success in bringing about some specified condition—in the classification of mental states. We are inclined to think, with Dennett, that the distinction between merely attributed intentionality and real intentionality can be understood within this framework, rather than constituting a refutation of it. From this perspective, we try to practice an approach that is an amalgam of the design-oriented approach of work in artificial intelligence with H.P. Grice’s conceptual creature construction.⁵ This approach seeks to understand intentionality through understanding the reasons that intentionalistic theories are useful in dealing with (or designing) various systems, beginning with very subhuman systems. We think that cognitive psychology depends on naturalistic psychology and that Fodor’s arguments against the possibility of the latter are not very convincing.⁶

From both our perspective and Fodor’s, it is necessary to have an account of how the contents and causal powers of the structured states of individuals intelligent beings can mesh in the way presupposed by common sense and by cognitive psychology. We try to provide such an account in this paper.

2.2 An Example

We shall look closely at an example that brings out the problems that concern us. Suppose Jerry wants to drink some decaffeinated coffee. He sees before him a brown cup, *c*; just a few seconds before, he had seen a brown cup being filled with what he believed was decaffeinated coffee. He moves his arm and hand in a complicated manner, grasping the cup, lifting it to his lips at an angle as he tilts his head a bit, and opens his mouth. Call this type of movement *M* for later reference. In Jerry’s circumstances, this movement constitutes picking up the cup and bringing it to his lips. Gravity and

⁴Our thinking on this point has been much influenced by conversations with Brian Smith. He addresses the issues concerning formality in **A View from Somewhere: Foundations of Computation and Intentionality**, in preparation.

⁵H. P. Grice, *Method in Philosophical Psychology (From the Banal to the Bizarre)*, **Proceedings and Addresses of the American Philosophical Association**, 1975.

⁶Some of these ideas are developed, or at least hinted at, in our paper, *What Is Information?*, to appear in **Information, Language and Cognition**, Vancouver Studies in Cognitive Science, Vol. I, ed. P. Hanson, Vancouver, University of British Columbia Press, 1989, in press.

his digestive system take over, and he gets what he wants. The explanation of Jerry's doings might go like this:

Why did Jerry decide to pick up the cup and bring it to his lips?

He wanted to drink some decaffeinated coffee, and he believed that the cup in front of him was filled with decaffeinated coffee.

This example can be used to illustrate the problems engendered by the fact that common sense psychology classifies mental states in terms of external, circumstantial denominations. We explain Jerry's action by citing a belief and a desire. The belief and the desire make sense of the action. The action will promote the satisfaction of the desire if the belief is true. But the way the belief, desire, and action are individuated appears to depend on factors external to the formal properties of the mental tokens involved. One way to see this is by focusing on what we count as having *the same belief*. Construed transparently, the belief attributed to Jerry is individuated in terms of a certain cup *c*, rather than the way Jerry thinks of it. Someone would agree with Jerry—would have the same belief—who believed that *c* contained decaffeinated coffee, no matter how the other person thought of *c*. Yet only the ways of thinking, not the external reference, could conceivably be correlated with the formal properties of belief tokens.

Construed opaquely, things are a bit better. But the belief still seems to be individuated in terms of Jerry.⁷ As Fodor points out, opaque attributions still allow some slack as to how the agent is thinking. Someone else, who is looking both at Jerry and at *c* but thinking “The cup in front of him contains decaffeinated coffee,” would be said to believe the same thing that Jerry does. But then it is at least possible that Jerry could believe just what he does, opaquely construed, while being in a different state.⁸ Perhaps, sitting in a Denny's, gazing in a mirror while waiting impatiently for his cup of decaf, Jerry sees a waitress fill a customer's cup from the pot with the characteristic decaf indicator—the pot has an orange neck. He thinks, “Now the cup in front of him contains decaffeinated coffee...I wonder when the cup in front of me will do so.” In such a circumstance, it would be misleading to report that Jerry believes the cup in front of him contains decaffeinated coffee; but would it be incorrect, would what the reporter said be false? It seems not, for as we noted, someone else, who thinks “The cup in front of that man contains decaffeinated coffee”, with reference to Jerry, would be counted as believing just what Jerry believes when Jerry thinks “The cup in front of me contains decaffeinated coffee.”⁹ So it seems that opaque attitude reports also do not

⁷Here, we follow Fodor's discussion of *transparent* and *opaque* classifications, in *Methodological Solipsism*. Note in particular his treatment of the “I'm sick”/“He's sick” and the “I'm ill” examples.

⁸Following Fodor and many others, we have allowed ourselves use of locutions of the form *A thinks (believes) “...”* We regard this as an alternative notation for fully opaque belief attributions, which means that there is a token of the quoted type in the agent's belief structure. See below, §4.2. It is interesting to note that this odd way of speaking is often accepted without much explanation.

⁹On this, compare Fodor's treatment of the “That's edible”/“This is edible” example in *Methodological Solipsism*.

focus exclusively on the sorts of intrinsic (formal) properties needed by a causal theory of mind.¹⁰

3 The Need for Circumstantial Content

Fodor uses the phrase “object of an attitude” to refer to the mentalese tokens that are involved in cognition according to his theory. We try to avoid this phrase, which is often used instead for what we shall call *the contents of the propositional attitudes*. We assume that the tokens have contents, and that these contents are referred to by the complements of attitude reports. So “*A* believes that *P*” means, roughly, that *A* has in his head a belief token, whose content is that *P*. We take these contents, the designata of phrases of the form “that *P*,” to be propositions. Propositions are truth-evaluable; that is, not true or false as spoken by one person, or at one time, but simply true or false.¹¹ We use “believes that ...,” “desires that...,” and “intends that...” as canonical forms for reporting beliefs, desires, and volitions.¹²

Sentences in ordinary language may be separated into those that are *eternal* and those that are *context-sensitive*. Different utterances of an eternal sentence express the same proposition. The class of eternal sentences is arguably empty, but sentences such as “ $7+5=12$ ” and “The first person born in the twenty-first century was, is, or will be a philosopher” are at least candidates.¹³ Different utterances of context-sensitive sentences express different propositions, depending on such factors of the utterance as the agent, the time, the persons the agent is addressing, the objects the agent is attending to, and other wider circumstances. The class of context-sensitive sentences includes those with indexicals, demonstratives, tense markers, and, arguably, proper names. Thus if Jerry and Zenon each say, “The cup in front of me contains coffee,” they express different propositions that may be true and false independently. Context-sensitive sentences in mentalese will work the same way; if Jerry and Zenon each have “The cup in front of me contains coffee” (= *S*) written in their belief structure, they will believe different things. But is there any reason to suppose that there are context-sensitive sentences in mentalese? What reasons would a psychological theory that

¹⁰On the view of belief reports developed in Jon Barwise & John Perry, *Situations and Attitudes* (Cambridge, MA, MIT/Bradford Books, 1983), this report would be misleading but literally true. On the view developed in Mark Crimmins and John Perry, *The Prince and the Phone Booth: Reporting Puzzling Beliefs*, *Journal of Philosophy*, forthcoming, there is a reading, indeed the natural reading, on which the report would be literally false. But even on that reading, the proposition that Jerry is reported to believe is the same as that he expresses with his use of “The cup in front of me contains decaffeinated coffee”—even though the *beliefs* are different. The present authors subscribe to the latter approach. Here, however, we are dealing with the propositional attitudes themselves, not the semantics of attitude reports. It is crucial for the strategy in *The Prince and the Phone Booth* that beliefs be concrete cognitive structures. It therefore fits well with Fodor’s token-oriented approach. In his dissertation, *Talk About Beliefs* (Stanford University, Philosophy Department, 1989), Crimmins discusses a variety of ways in which attitudes could be concrete cognitive structures that are less language-like than mentalese is often taken to be.

¹¹Relativity to a possible world is a different matter.

¹²For present purposes, we are simply identifying volitions and *intentions in action*.

¹³But the first seems relative to a base—it’s true base 10, false base 12—while the second depends on a frame of reference for fixing dates, which seems to smuggle in some noneternality.

adopted Fodor's language of thought hypothesis have to include such sentences in the language postulated?

Suppose Jerry and Zenon are each thirsty, and each has a cup of coffee in front of him. They each execute movement M , and each satisfies his desire. Both the similarities and differences are easy to conceive, if we suppose that the beliefs Jerry and Zenon have are context-sensitive—in particular, if what they believe, in virtue of the occurrence of a token of S in their heads, is systematically related to the same simple fact about that token, namely in whose head it does occur. For instance, we can imagine that Jerry and Zenon have different instances of a certain type V of visual impression;¹⁴ that because of these impressions a token of S is written in the belief structure of each; and that this, given the presence of “I drink coffee” in the desire structure of each, causes a token of “I pick up the cup in front of me and drink from it,” to be written in their volitional structure, which causes each of them to execute M . Jerry and Zenon thus go through the same succession of states, and these processes are instances of the same psychological law. But Jerry and Zenon see, believe, desire, intend, and do different things. In particular, what Jerry believes could be true, while what Zenon believes is false, and vice versa. What Jerry believes (that there is a cup in front of Jerry) would not be a good reason for Zenon to execute M . It might be true that there is a cup in front of Jerry, while there is an irritable gorilla in front of Zenon. Executing M in these circumstances would not be a good idea.

We want to emphasize that it is also important to have beliefs that are less sensitive to context. Consider the case with which we began. Jerry doesn't just want coffee; he wants decaffeinated coffee. A moment or two ago he saw that a certain cup contained decaffeinated coffee. Now suppose that since that time he has left the room; the cup became perceptually inaccessible to him. He needs to be able to store the relevant information about the cup, in a way that retains the same content through the change in context, as he diverts his attention from the cup and leaves the room, and throughout the [short] interval of his absence. And he needs to retain it in a form that will allow him to reidentify the cup upon his return and make use of his knowledge. This requires a way of thinking of the cup that continues to pick out that particular cup through the changes of context. This way of thinking will probably be context-sensitive, but not as context-sensitive as the way of thinking of the cup that is associated with a contemporaneous visual perception of it. For example, he might think of it as “the cup I saw at my table a moment ago.” This way of thinking does not shift its reference as Jerry moves about or shifts his attention to other objects.

A (relatively) basic physical action, like the execution of M , has different results depending on who does it, when, and in what circumstances. A system of beliefs should lead to the actions that will be successful, relative to one's desires, in the circumstances that make the beliefs true. But humans do not just depend on their perceptions at the time of action to provide them with the beliefs needed for decision and volition, but earlier perceptions, inferences, and the perceptions and inferences of others. In all

¹⁴We are supposing then that they are similarly situated with respect to their respective cups, in so far as their perceptual apparatus is concerned.

of these cases in which the pragmatic effect of information is delayed or distanced, ways of preserving content through change of context are needed. It would be very misguided to take the simple and relatively automatic case of picking up a cup of coffee when thirsty as a complete guide to the role of the attitudes. Still, the need for context-sensitive representation never goes away. If he wants to drink the coffee, Jerry still has to pick up the cup in front of him.

Consider, for a different example, Jerry's making a phone call to Zenon. Here the belief, say "Zenon's phone number is 555-5555," is relatively context insensitive. There is the present tense marker, but phone numbers are relatively stable properties of persons, so we can ignore that. When Jerry calls Zenon, the perceptions that originally gave rise to the belief may be remote in time, and the effect of the action that fulfills the goal—the ringing of Zenon's telephone and his answering it—are remote in space. Neither the number nor Zenon needs to be thought of demonstratively or indexically to understand the transaction. But, like all human action—except perhaps pure ratiocination—the crux of the matter involves physical interactions with a physical object. The practical reasoning involved will terminate with some context-sensitive way of thinking about the telephone. The Fodorian version of half of Kant's maxim about concepts and intuitions should be "Eternal tokens of mentalese without context-sensitive tokens of mentalese are blind."

4 The Basic Concepts

4.1 A Fodorian Model of Cognition

We have so far been supposing a Fodorian model; we shall now sketch an extremely simple version of a mentalese account of the structure of cognitive states. We note that states are repeatable types; we shall speak of instances of such. Instances are concrete, nonrepeatable episodes. We shall assume that there are three functionally specified components of the mental states of our agent: the belief component, the appetitive component, and the volitional component. These three components are realized in three distinct, but connected, concrete structures. Conceived of abstractly, that is, functionally, each of these is like a file, into which tokens of sentences are read, in which they can be stored and manipulated in various ways, and from which they can be read. We shall further assume that mentalese is a lot like English; indeed, we shall assume that it is English.¹⁵ With reference to the first example, we might expect to find the following in Jerry's mind:

- In the belief structure, a token of "The cup in front of me has decaffeinated coffee in it."
- In the appetitive structure, a token of "I drink some decaffeinated coffee."
- In the volitional structure, a token of "I pick up the cup and bring it to my lips."

¹⁵We hope it is clear that this assumption is for the sake of simplicity, not chauvinsim.

These tokens are concrete structures, characterizable in many different ways. We have classified them by their syntactic type, given that they are tokens of English. This leaves open many difficult questions about the relations between type and token; we shall ignore these here. We can classify agents in many different ways, too, of course. We introduce three relational symbols, $thinks_B$, $thinks_D$, and $thinks_V$ for the relations that hold of an agent A , a time t , and a type T , just in case there is a token of type T in A 's belief (appetitive, volitional) structure at t . We shall usually ignore the temporal relatum. We shall also treat these symbols as if they were transitive verbs. Thus $thinking_D$ "I drink some decaffeinated coffee" is an appetitive state and is also a (partial) mental state. Many different agents, at different times and locations, can be in that mental state, with or without being in the $thinks_B$ state we supposed Jerry to be in. Of course, no two instances of these states are identical; we leave open the possibility that a single agent, at a single time, can have two distinct tokens of the same type in one of its mental structures.

Following Fodor, we shall assume that there is a single central processor, which can read the tokens in the various structures and can perform various operations on them. The processor is a deterministic device. In this respect, then, the agent's psychology is lawlike. We further assume that every agent of a given kind or species has the same kind of central processor, and indeed this is partly definitive of what we mean by a kind or species of agent. Of course, different experiences will have led to quite different sentences being written in the structures of various agents of the same species.

As noted, the psychologies of our agents are law-like. There are laws relating the various states. Here is a candidate law, implausibly simple, that we will assume to hold:

(\mathcal{L}): If an agent $thinks_B$ "The cup in front of me has decaffeinated coffee in it," and $thinks_D$ "I drink some decaffeinated coffee," it will (normally or *ceteris paribus*) come to $think_V$ "I pick up the cup and bring it to my lips."

The above statement of the law conceals reference to tokens of the displayed types, but for there to be laws like \mathcal{L} , or like \mathcal{L} except for being much more complex, the syntactic type of a mentalese token must be a property of tokens that the processor can detect. We assume that the processor can only detect local, physical properties of the tokens. It is not quite clear what this includes, but there are a number of things that are clearly not included. For example, the processor cannot detect anything about the cup or the coffee. This is not to say that the agent's sensors can't detect such things; nor is it to say that the agent can't. In any event the law, as stated, is quite independent of the meaningfulness of the tokens and of whatever particular contents they have.

We need to make three points about \mathcal{L} and the extremely simple psychology it reflects. First, we are ignoring background beliefs and a number of interesting questions they raise. The use of the phrase "normally or *ceteris paribus*" is simply an indication that we are aware of these issues, not an attempt to treat them.¹⁶ Second, we are

¹⁶Also, see below §5. We basically agree with Fodor (**Psychosemantics**, pp. 4ff.) that the laws of psychology will be *ceteris paribus* laws. But we suspect that physical laws are as dependent on *ceteris*

making no allowance whatsoever for weighing the pros and cons of various alternative courses of action or for deliberation to resolve conflicts among appetites. Finally, we are using the volitional structure to model central motor control functions. Belief states and appetitive states lead to volitional states; these, in turn, cause bodily movements. A given volitional structure could be wired up to the wrong kind of body, one that had nothing like hands with opposable digits, or to a body with arms that were too short, etc. It could also be ill-wired to the right kind of body. We ignore all such unhappy possibilities.

4.2 A Semantics for Mentalese

We assume that a cognitive psychology assigns contents to tokens of mentalese in virtue of (i) the basic meanings associated with the types of the tokens and (ii) other facts. These other facts we gather into the following basic categories:

- (A) Facts about the tokens themselves, specifically whose head they occur in (or, as we shall say, who owns them), and when; these facts determine the reference of such indexicals as “I,” “me,” and “now.”
- (B) Facts about the owners of the tokens, such as which objects they are attending to, talking to, and the like. These facts, together with facts of category A, determine the reference of indexicals and demonstratives such as “you,” “he,” “her,” and the like.
- (C) Other facts that may be relevant, together with facts of categories A and B, in determining the reference of definite descriptions and perhaps names.¹⁷ This may include facts that are not in any clear and intuitive sense about either the token itself or its owner.

We take the meanings associated with the types to be functions from circumstances to contents, where contents may have as constituents individuals external to the mind of the agent. In these respects, our semantics follows the semantics for “schemata” in Chapter Ten of **Situations and Attitudes**. But we differ with that approach in two related ways. First, we do not assign a single content to a token. Each token will have three contents: a fully opaque content, an opaque content, and a transparent content. Second, we do not take the function from context to content(s) to be the basic fact about meaning, but to result from a basic assignment of *token reflexive conditions of truth*. Thus our semantics will assign (up to) three contents to each expression: truth conditions—which we will take to be the fully opaque content—opaque, and transparent content. All are assigned circumstantially.

paribus clauses as any others. See Nancy Cartwright, **How the Laws of Physics Lie**, Oxford University Press, 1986.

¹⁷Fodor’s remarks on transparency, opacity, and full opacity are not too definitive with respect to names, and we steer clear of them in this essay.

The following is an attempt to indicate the form of our semantic account by treating the sentence “The cup in front of me contains decaffeinated coffee” in some detail. We start with the content of the terms “me” and “the cup in front of me”:

$T = \text{“me”}$

1. Basic condition of reference for a token t of $T = \textit{being the owner of } t$.
2. Opaque content of t (condition of reference, given that Jerry owns t) = *being [identical to] Jerry*.
3. Transparent content of t (condition of reference given that Jerry owns t and any other facts) is the same as the opaque condition = *being Jerry*.

$T = \text{“the cup in front of me”}$

1. Basic condition of reference for a token t of $T = \textit{being the unique cup in front of the owner of } t$.
2. Opaque content of t (condition of reference, given that Jerry owns t) = *being the unique cup in front of Jerry*.
3. Transparent content of t (condition of reference, given that Jerry owns t and that c is the cup in front of Jerry) = *being } c.*

$T = \text{“The cup in front of me contains decaffeinated coffee.”}$

1. Basic condition of truth for a token t of $T = \textit{Someone } x \textit{ is the owner of } t, \textit{ something } y \textit{ is the unique cup in front of } x, \textit{ and } y \textit{ contains decaffeinated coffee}$.
2. Opaque content of t (condition of truth, given that Jerry is the owner of t) = *Something } y \textit{ is the unique cup in front of Jerry, and } y \textit{ contains decaffeinated coffee}.*
3. Transparent content of t (condition of truth, given that c is the cup in front of Jerry) = *c contains decaffeinated coffee*.

To get at the level 1 and 2 contents (fully opaque and opaque), we shall use the sentences “The cup in front of the owner of t contains decaffeinated coffee” and “The cup in front of Jerry contains decaffeinated coffee.” When we do this we will be using the descriptions *attributively*. Thus in the first sentence, t is referred to, but the owner of t and the cup are not referred to. In the second, Jerry is referred to, but the cup is not.

We use the notions of *loading* and *unloading* to get at relations between the fully opaque and the opaque, and the opaque and the transparent contents of a token of T . Loading is an operation that takes us from a proposition that contains a complex property, like being the owner of t , to a proposition that contains an object that uniquely

instantiates the property, like Jerry. More precisely, we load a proposition with respect to a complex property and a set of circumstances. In going from the level 1 to level 2, we are loading the proposition with respect to the property of being the unique owner of t and the circumstance that Jerry is that owner. In going from the level 2 to level 3, we are loading the 2-level proposition with respect to the property of being the unique cup in front of Jerry, and the circumstance that c is that cup. Unloading is just the opposite of loading. The proposition that the cup in front of Jerry contains coffee is the result of unloading the proposition that c contains coffee with the circumstance that c is the cup in front of Jerry.¹⁸

We take the truth conditions of a mentalese token to be its fully opaque content. This is the only content it has that depends only on the form of the token and the interpretive function.

We said at the beginning of this essay, “Fodor takes propositional attitudes to be relations to tokens of an internal language...that have content. If Jerry believes that S , Jerry has a token of mentalese in his belief structure that has the content that S .” The picture we have arrived at requires an account that is a bit more complicated, however. First, we should note that when the context of the attitude reporter and the agent differ, different sentences will be required to get at the same content. Thus to report what Jerry believes in virtue of his having a token of “I am sitting” in his belief structure, I’ll have to say something like “Jerry believes that he is sitting” or “The author of *Methodological Solipsism* believes that he is sitting.” If I use a token of the same sentence, and say “Jerry believes that I am sitting,” I do not convey the right message.

Second, we now have three levels of content. We shall introduce subscripts and say things like

Jerry *believes*_{FO} (*desires*_{FO}) that...

Jerry *believes*_O (*intends*_O) that...

Jerry *believes*_T (*desires*_T) that...

This notation is not really adequate, since there are cases in which one term in a content sentence should be taken opaquely and another transparently. For the purposes of this essay, however, this notation will suffice.

We should emphasize that *believes*_{FO} is *not* one of our ordinary belief concepts. In the ordinary senses (opaque and transparent) of “believes,” people don’t usually believe the fully opaque content of their beliefs, since most people do not have beliefs about tokens of mentalese in their heads. And note further that even for people who do have such beliefs, the belief in the fully opaque content of a belief token b will not be the belief one has in virtue of having b . To see this, note that Jerry and Georges might both believe, of Jerry’s belief token b , that the cup in front of the owner of it contains decaffeinated coffee. Imagine Jerry and Georges are talking about mentalese,

¹⁸In *What Is Information?*, we sketch a theory of propositions, within a version of situation theory, in which these notions can be given precise, formal embodiments. Notice that the operation of unloading is *not* analogous to the allegedly impossible operation of going from a denotation (or reference) to a sense. The circumstantial fact, together with the object, determines the complex, typically relational, property.

using one of Jerry’s beliefs as an example, while Jerry drinks decaffeinated coffee. This belief about Jerry’s token is not the same belief that Jerry has in virtue of having the token. This may seem a bit puzzling. But suppose that Jerry uttered “The cup in front of me contains decaffeinated coffee.” Call the utterance u . The belief about u , that the cup in front of the person who made it contains decaffeinated coffee, has to be distinguished from what Jerry said (the proposition he expressed) with u . What Jerry said could be true, even if he never spoke, and u never existed. This is analagous to the belief case, and perhaps will make it seem less puzzling. Our ordinary propositional attitude reports simply do not focus on fully opaque content.

So we have to be careful with $believes_{FO}$ and remember that its meaning derives from the theory, not from common usage. Saying that Jerry $believes_{FO}$ that the cup in front of the owner of b contains decaffeinated coffee is just saying that b is a token in Jerry’s head, whose fully opaque content is that the cup in front of its owner contains decaffeinated coffee. Jerry doesn’t believe this in the ordinary sense. That is, he doesn’t $believe_O$ it. Georges does $believe_O$ it, but doesn’t $believe_{FO}$ it. So, of Georges, but not of Jerry, it can be said that he believes that the cup in front of the owner of b contains decaf.

5 Rational Laws and Adequate Explanations

The basic idea of a rational law is this. Suppose a belief and a desire cause an action. Then the action should promote the satisfaction of the desire, given the truth of the belief. This is a version of Fodor’s point in the initial quotation, which we have made before with respect to our example. Jerry’s belief that the cup in front of him contains decaffeinated coffee and his desire to have decaffeinated coffee lead him to will to move in a way that results, if the belief is true, in his desire being satisfied. There is surely something quite appropriate about this. Suppose Jerry’s belief and desire led him instead to a movement that results in splashing the coffee into his forehead. That would in some sense be inappropriate. As Fodor points out, it is perfectly conceivable that beliefs and desires, conceived as internal states, should have such inappropriate effects, but such effects of beliefs and desires are not what cognitive psychology is all about—they are not grist for the cognitivist’s mill.

It will help discussion to define a relation among four propositions P , R , Q and C , thought of as the contents of a belief, a volition, and a desire, and background conditions: (the belief that) P *rationalizes* (bringing it about that) R , relative to (the desire that) Q , given (the condition that) C . We’ll write this

$$Rationalizes_Q(P, R | C)$$

We think of this relation as one of *incremental rationality*.

If C is given, then if P is true also, bringing it about that R will guarantee (or at least promote) bringing it about that Q .¹⁹

¹⁹Since P , Q , etc. are propositions, and since we assume that the whole clause “that S ” (where S is a

To see one source of the need for, and role of, the background condition, consider the fully opaque contents of the tokens in our basic example. Let P = the proposition that the cup in front of the owner of b contains decaffeinated coffee, Q = that the owner of d (a token of “I drink decaffeinated coffee”) drinks decaffeinated coffee, and R = that the owner of v picks up the cup in front of him and brings it to his lips. Then it is not the case that bringing it about that R is true will bring it about that Q is true, if P is true—*without the further condition that b , d , and v belong to the same agent*. This is our reason for saying that without some appeal to circumstances, the rationality of laws of cognitive psychology cannot be understood.²⁰

Now consider a law, like our \mathcal{L} , to the effect that a belief token b of type B and a desire d of type D , in the same agent, will cause that agent to have a volition v of type V . Such a law is *rational* if for every b , d , and v that instantiate it,

$$\text{Rationalizes}_{\text{Con}_{FO}(d)}(\text{Con}_{FO}(b), \text{Con}_{FO}(v) \mid C)$$

where $\text{Con}_{FO}(b)$ is the fully opaque content of b , and where C is the condition that b , d , and v belong to the same agent (and conditions are normal). When beliefs and desires cause volitions in accord with such laws, we can say that they motivate in two senses. The belief and desire cause the volition, and they rationalize it.

We should say a word about *normal conditions*. In our conception of cognitive psychology, the specification of an environment, which includes specification of a range of normal conditions, will be an important part of the psychological theory for a kind or species of agent. In our conception, agents are attuned to certain environments, and the apparatus of perception and belief is used to pick up and store information about factors that vary within those limits. We see the circumstantial nature of thought as one aspect of this attunement. The way visual information is used to guide our hands, for example, involves attunement to the normal relations between the orientation of limbs and the orientation of eyes. Such attunement only goes so far, however, and that is where the function of systems of storeable, manipulable, relatively context insensitive representations come to the fore—for instance, as they are involved in belief. These points are rather tangential to our main aim in this essay, however.

The conception of a rational law explained a few paragraphs earlier makes it clear that fully opaque explanations, involving rational laws, would be fine. By subsuming behavior under causal laws, they would also subsume it under rational laws. But as we mentioned, we don’t ordinarily use attitude reports to attribute fully opaque belief—what we say when we say “Jerry believes that...” is not understood on the model “Jerry *believes*_{FO} that...” We want then to consider how explanations using ordinary opaque and transparent attitude attributions could be adequate, and how they can go wrong.

sentence) refers to a proposition, consistency would dictate saying “the belief S ” rather than “the belief that S .” But it sounds better the inconsistent way.

²⁰This should not be taken to exclude the necessity of other background conditions as well, such as that the movement is taking place in the Earth’s gravitational field. Some background conditions might themselves be the contents of the agent’s background beliefs; the condition just noted is not likely to be one of those. See below.

The basic idea is as follows. When we explain an action in terms of a belief and a desire, we are basically explaining the occurrence of a volition of a sort for which the action is a more or less basic mode of execution.²¹ The explanation will be correct only if it can be unloaded down to an instance of a rational law where the consequent volition is so executable. To see what this means, let's work through our example.

Consider this explanation of Jerry's executing movement M , thereby picking up c and bringing it to his lips:

Jerry believes c contains decaffeinated coffee and desires that he drink some decaffeinated coffee.

What does this come to? Here's what:

- Jerry *believes_T* that c contains decaffeinated coffee in virtue of *believing_O* that the cup in front of him contains decaffeinated coffee together with the external circumstance that c is the cup in front of him. He *believes_O* this in virtue of owning a belief b whose fully opaque content is that the cup in front of the owner of b contains decaffeinated coffee.
- Jerry *desires_O* that he drink some decaffeinated coffee in virtue of owning a desire d whose fully opaque content is that the owner of d drink some decaffeinated coffee.
- Movement M is a basic way of executing a volition v whose fully opaque content is that the owner of v pick up the cup of coffee in front of him.
- Given the assumption that there is a rational law like \mathcal{L} above, the belief and desire cited explain the action.

Now consider the case in which Jerry looks in the mirror. Here he doesn't actually perform any action. But we want to understand why, knowing about his belief and desire, we shouldn't expect an action such as his reaching for the cup by executing movement M .

The type of the volition and the desire are just as before. Further, Jerry again *believes_T* that c contains decaffeinated coffee in virtue of *believing_O* that the cup in front of him contains decaffeinated coffee, in the circumstance in which c is the cup in front of him. He *believes_O* this, however, in virtue of owning a token b' with the fully opaque content that the cup in front of the man whom the owner of b' is watching contains decaffeinated coffee. The belief b' , the desire d , and a volition v of which the (missing) action would be an execution are not instances of a rational law. They are not instances of \mathcal{L} , because the fully opaque content of the belief is not right. Moreover, if there were a law linking beliefs of this type, with this fully opaque content, with desires and volitions of the types of d and v , it would not be rational. Lucky folks like Jerry

²¹The route from action to volition will be more complex when the the context of explanation includes background about other beliefs and desires of the agent that could be involved in more complex strategies of execution.

would get a drink of coffee, but in most cases the agent would be knocking over things, pawing empty space, or irritating irritable gorillas.