

## Solipsistic Semantics

In a famous passage of the *Meditations*, Descartes writes:

At this moment it does indeed seem to me that it is with eyes awake that I am looking at this paper; that this head which I move is not asleep, that it is deliberately and of set purpose that I extend my hand and perceive it . . . . But in thinking over this I remind myself that on many occasions I have been deceived by similar illusions, and in dwelling on this reflection I see so manifestly that there are no certain indications by which we may clearly distinguish wakefulness from sleep that I am lost in astonishment. And my astonishment is such that it is almost capable of persuading me that I now dream.<sup>1</sup>

In his skeptical arguments, Descartes is claiming not merely that it is possible that all his thoughts about the world are false but that it is possible for him to have these very thoughts and that they be false. It is possible that he is dreaming that there is a world outside his mind though none exists. Descartes's radical skepticism also involves his view that he can know the contents of his thoughts, even though he knows nothing of the world, or knows even if there is a world.

On the Cartesian picture, the content of a thought is a property intrinsic to the thought and conceptually independent of any individuals outside the mind.<sup>2</sup>

Semantics for a system of mental representation assigns to each mental representation a meaning or content. The assignment can take a number of forms, for example, a truth theory, an assignment of Fregean senses, or even an image theory. We will say that semantics is solipsistic (hereafter *SS* for "solipsistic semantics") if in assigning meanings to representations it does not presuppose the existence of any mental or physical individuals other than the thinker and his thoughts. This characterization needs to be sharpened. There are at least two ways in which an assignment of meanings to representations may presuppose the existence of individuals. One is for the semantics to *interpret* a representation by assigning as its meaning some individual other than the thinker and his representations. For example, semantics that interprets a proper name as directly referring to an actually existing individual implies that the reference of the name exists or did exist. The other sort of presupposition is a bit more difficult to specify.

Given a language *L* and an interpretation for *L*, we can ask the question, In virtue of what does *L* have that interpretation? For an interpretation of mental representations, the question is, What is it about the representations, their structures and other intrinsic properties, their interactions with each other, with the thinker's physical and social environment, and so forth, in virtue of which they have that interpretation? We will call a theory that answers this question for

a language  $L$  with interpretation  $I$  a *theory of meaning* for  $L$ . It is important to distinguish the question a theory of meaning is supposed to answer from the question of what events happen to cause a representation to have a particular interpretation. The distinction is parallel to two ways of understanding the question, What makes a good man good? One question is, What causes a man to be good? (or how can we make a man good?). The other question is, What kinds of facts make a good man good? The second question is a conceptual or metaphysical question about the nature of goodness. It is this kind of question concerning meaning that a theory of meaning attempts to answer.

An example of a theory of meaning that does not presuppose the existence of individuals other than the thinker and his thoughts are certain "picture" or "image" theories of meaning.<sup>3</sup> On this view, thoughts are mental images whose representational powers are determined by intrinsic features of the image, for example, its phenomenal color and shape. The image refers to whatever resembles it, but it has the meaning it does entirely in virtue of its intrinsic features. An example of a theory of meaning that does presuppose the existence of individuals external to the thinker is Kripke's causal theory of names. A name means what it does in virtue of bearing a certain causal relation to its bearer. As we pointed out, Kripke's interpretation is also nonsolipsistic since it interprets a name as meaning its bearer. It may be that the correct theory of meaning for some kinds of expressions is solipsistic whereas the correct theory of meaning

for other kinds is nonsolipsistic. It is plausible that the mental counterparts of logical constants possess the meanings they do in virtue of their functional roles and that functional role is solipsistic. *SS* for *L* requires that neither the interpretation nor the theory of meaning for *L* presuppose the existence of individuals external to the thinker.

A theory of meaning and interpretation for *L* may be solipsistic even though *L* contains terms that *purport* to refer to individuals external to the thinker. This is just the sort of possibility Descartes envisaged. The central idea of *SS* is that the determinants of the meanings of one's mental representations are entirely within oneself. A doctrine closely related to *SS* is that if two of a person's mental representations have the same or different meanings, then it will be possible for him to determine that they have the same or different meanings by introspection alone. We will call this doctrine "transparency". Given the accessibility of one's mental representations to consciousness, transparency is entailed by there being *SS* for mental representations.

In this paper we will examine Cartesian and more contemporary motivations for the view that there must be *SS* for mental representations. We will discuss some well known arguments that show that various kinds of natural language expressions can possess nonsolipsistic semantics. An apparent consequence of this is that when sentences containing such expressions are used to specify thoughts, as in attributions of propositional attitudes, the thoughts are

characterized nonsolipsistically. We argue that semantics for English is so thoroughly nonsolipsistic that even if thoughts have *SS* their contents cannot be expressed in English. We then argue that the most plausible theories of meaning for mental representations are also nonsolipsistic. Our discussion results in an apparent dilemma. On the one hand, there are the Cartesian intuitions and other motivations for thinking that thoughts possess *SS*. On the other hand, there are the arguments that seem to show that given our usual ways of characterizing thought contents, their semantics is nonsolipsistic. We conclude with some tentative remarks on how these two views might be reconciled.

Descartes is not alone in his advocacy of *SS*. Hume remarks that "to form the idea of an object and to form an idea 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."<sup>4</sup> Hume is saying that the interpretation of his ideas is solipsistic since one can have an idea of an object even though the object fails to exist. Hume's image theory of meaning is also solipsistic since it locates the representational powers of an idea in its intrinsic features. Frege also seems to endorse *SS*.

Let us just imagine that we have convinced ourselves, contrary to our former opinion, that the name *Odysseus*, as it occurs in the *Odyssey* does designate a man after all. Would this mean that sentences containing the name "*Odysseus*" expressed different thoughts? I think not. The

thoughts would strictly remain the same; they would only be transposed from the realm of fiction to that of truth. So the object designated by a proper name seems to be quite inessential to the thought-content of a sentence which contains it.<sup>5</sup>

Frege is saying that the name "Odysseus" has the sense it has whether or not Odysseus exists. His account is that to understand the name is to grasp its sense. Although he doesn't say much about what it is to grasp a sense, his view seems to be that the grasping of senses is a matter strictly between the mind and the realm of senses. It requires the existence of no physical or mental individuals (senses are neither mental nor physical) other than the thinker and his thoughts. Frege also holds "transparency." He frequently uses it to establish that two expressions have different senses. Since one can grasp the senses of "Hesperus" and "Phosphorus" without realizing that they have the same reference, it follows, according to Frege, that they have different senses.

Descartes and Frege were dualists. On their accounts, the mind possesses an intrinsic and unexplained power to represent the world. There are also physicalist versions of *SS*. Physicalism will attempt to account for semantic facts in terms of physical properties and laws. A physicalist who endorses *SS* for mental representations thinks that the physical facts that determine the meanings of thoughts involve only intrinsic physical properties of the thinker's body. One can

express this view with the claim that the semantic properties of a thinker's mental representations *supervene* on intrinsic physical (e.g., neurophysiological) states of his body.<sup>6</sup>

Physicalistic *SS* can make for strange bedfellows. Two prominent contemporary proponents are John Searle and Jerry Fodor, who agree on little else in the philosophy of psychology.<sup>7</sup>

Searle expresses *SS* as follows:

If I were a brain in a vat I could have exactly the same mental states I have now; it's just that most of them would be false... The operation of the brain is causally sufficient for intentionality. It is the operation of the brain and not the impact of the outside world that matters for the content of our internal states. [BBS 452]

I think in the relevant sense that meanings are precisely in the head there is nowhere else for them to be. [Int 200]

Searle's view is that the causal operations of the brain are sufficient (and perhaps necessary) to produce thoughts with their contents.

Fodor's version of *SS* is different from Searle's. Fodor advocates (whereas Searle rejects) the "computational theory of mind" (hereafter, *CTM*). According to Fodor, mental processes are

analogous to the operations of a computer. For example, when one forms an intention to, say, go to a certain Chinese restaurant for dinner, one's mind (brain) engages in computational processes involving the manipulations of various mental representations. Fodor argues how a great deal of theorizing in cognitive psychology presupposes *CTM*. He also argues that *CTM* provides plausible explanations of various features of mental states and processes (e.g., the opacity of belief). Fodor argues that *CTM* is committed to the following "formality condition":<sup>8</sup> if psychological states (processes, etc.) have the same computational characteristics, then they must be the same psychological state (process, etc.). Fodor also maintains that psychological states (or an important subset of them) are characterized in terms of their contents. It follows from this and the formality condition that, if states are computationally the same, then they have the same contents (or rather are composed of representations with the same contents). In other words, thought contents supervene on computational features of thought. Since computational features, whatever they might be, supervene on intrinsic physical features, it follows that thought contents supervene on physical features. In this way Fodor seems committed to there being *SS* for mental representations.

Fodor's view is more complicated than we have so far indicated. Referring to considerations that we will soon discuss, he observes that our normal attributions of belief violate the formality condition. But he thinks that they come close to satisfying it. He writes that

"taxonomy with respect to content may be compatible with the formality conditions plus or minus a bit."<sup>9</sup> The suggestion is that even if our usual scheme of belief attributions assigns contents nonsolipsistically, it is close enough so that with a little tinkering, we can construct *SS* for mental representations.

Fodor's reasons for thinking that there exists *SS* for mental representations are mainly theoretical. He holds that modern cognitive theory requires it. *SS* also has a powerful intuitive appeal. The intuitions that underlie *SS* are supported by the Cartesian thought experiment. One can imagine each of his thoughts about the external world being false, even the thought that there is an external world, whereas the thoughts themselves remain the same. So it might seem that thoughts have the contents they do independently of any individuals that are extrinsic to them. Furthermore, although one might be mistaken about the meanings of words in one's public language, it seems absurd to think that one might be mistaken about the meanings of one's own thoughts. Descartes beautifully expresses this line of thought.

Now ideas considered in themselves and not referred to something else, cannot strictly be false; whether I imagine a she-goat or a chimera, it is not less true that I imagine one than the other ....

The chief and commonest error that is to be found in this field consists in my taking ideas within myself to have similarity or conformity to some external object; for if I were to consider them as

mere modes of my own consciousness, and did not refer them to anything else, they could give me hardly any occasion of error.<sup>10</sup>

Despite the cognitivist's arguments and the Cartesian's intuitions, the very possibility of *SS* is called into question by recent developments in the philosophy of language and mind associated primarily with work by Kripke, Kaplan, Putnam, and Bruegel. Taken as a whole, this work apparently shows the inadequacy of the Fregean account of meaning. Because of the close connection between Frege's theory and *SS*, these developments have a bearing on the possibility of *SS* for thought.

Saul Kripke's work on names was the spearhead of the attack against the Fregean accounts.<sup>11</sup> According to Frege, a name - for example, "Aristotle" - expresses a sense that, as it happens, picks out a certain referent (Aristotle). If things had turned out differently, the same sense might have picked out someone else or nothing at all. Kripke argues that Frege is wrong. A name does not express a sense but, instead, *directly* refers to its bearer. In terms of the apparatus of possible world semantics, a name refers to its bearer at every possible world. A term that expresses a sense - for example, a definite description - may refer to different individuals at different worlds. According to Kripke, the reference of a use of a name is determined by a causal chain that begins with a baptism of the bearer with the name. Someone

who hears the name "Aristotle" used by a competent speaker can himself use it to refer to Aristotle even if the information that he associates with "Aristotle" is insufficient to determine Aristotle (e.g., he is a Greek philosopher) or even if the information is uniquely true of someone else (e.g., he is the greatest Greek playwright).

How is Kripke's account of the meaning of names relevant to the semantics for mental representations? It is natural to assume that when someone is reported as thinking the thought that Aristotle was wise, he is said to have a thought whose content is the same as the content of "Aristotle is wise." His mental representation contains a constituent that directly refers to Aristotle. These semantics are nonsolipsistic in both ways discussed earlier. The interpretation of "Aristotle" is Aristotle himself and so involves an individual other than the thinker and his thoughts. Also, Kripke's account of what determines the interpretation of a name requires that its bearer exist at the time of baptism. So Kripke's theory of meaning for names is nonsolipsistic. We also note that "transparency" fails for Kripke's semantics. Someone might use the names "Hesperus" and "Phosphorus" which, according to Kripke, have the same interpretation, and yet have no way of discovering by introspection that they have the same interpretation. Meaning, according to Kripke, is not entirely in the head.

Kripke's direct reference account of names is superior to Frege's sense theory in a number of ways. However, there are problems with the account that seem to support the view

that names also must have a solipsistic interpretation that is relevant when they are used in contexts ascribing thoughts. If names are rigid designators, then since "Tully" and "Cicero" designate the same man, the thought that Tully was an orator and the thought that Cicero was an orator are thoughts with the same interpretation. But it seems possible to believe one without believing the other.<sup>12</sup> If believing that  $p$  is to be explained (as in Fodor) as tokening a mental representation that means that  $p$ , then it seems that the mental representations corresponding to "Tully was an orator" and "Cicero was an orator" must have different interpretations.<sup>13</sup> Another perhaps even more serious problem is that it certainly seems possible to think that Homer was a Greek even though it turns out that Homer never existed. But on the direct reference account, if Homer never existed, "Homer was a Greek" would fail to express any proposition.<sup>14</sup> It was precisely consideration of these problems that led Frege to postulate senses as the meanings of names. It is essential to his solution to the problems that whether a sense determines a reference or whether two senses determine the same reference is irrelevant to the grasping of senses. This suggests that semantics for names that is adequate for propositional attitude contexts will also interpret them as expressing senses.

David Kaplan has proposed semantics for indexical sentences that is nonsolipsistic.<sup>15</sup> According to Kaplan, when Arabella utters the sentence "She is a spy," pointing at Barbarella, she asserts a proposition that is essentially about Barbarella. This proposition actually contains

Barbarella as a constituent. If it turns out that Arabella is not pointing at anyone, then, on Kaplan's view, her utterance simply fails to express a proposition. It is clear that Kaplan's semantics interprets indexical utterances nonsolipsistically.

Suppose that when Arabella utters "She is a spy," she expresses a thought that has the same interpretation as her utterance. On Kaplan's account, this thought would contain a constituent that directly refers to Barbarella. According to this account, if Arabella and her neurophysiological twin point respectively at Barbarella and Twin Barbarella and each utters "She is a spy," they are thinking different thoughts. So the interpretations of mental representations containing indexicals do not supervene on neurophysiological states.<sup>16</sup>

Even if indexical thoughts have nonsolipsistic interpretations, there also seems to be a need to associate solipsistic interpretations with them. From Arabella's point of view, she is thinking the same thought when she utters "She is a spy," whether or not she is pointing at anyone. If thoughts are individuated in terms of their causal consequences for behavior, then it makes no difference whether or not Arabella is pointing at Barbarella, her twin, or is just hallucinating. Kaplan's distinction between the character of an indexical sentence and the proposition expressed by an utterance of the sentence may provide the ingredients for a solipsistic interpretation of indexical thoughts. Character is a function from contexts of utterance to propositions. For example, the character of "I am in Ann Arbor" uttered by Arabella at time  $t$

yields the proposition that Arabella is in Ann Arbor at  $t$ . This proposition contains Arabella herself as well as Ann Arbor and  $t$  as constituents. Semantics that assigns this proposition to indexical thoughts is clearly nonsolipsistic. But semantics that assigns to the thought its character might be compatible with SS. Arabella can have the character of "She is a spy" in mind, even though she is pointing to no one. Arabella and her twin have the same character in mind.

Character is not sufficiently robust to characterize the semantics of mental representation in all cases.<sup>17</sup> This can be seen in the following situation. Arabella is looking at two TV screens. One shows Barbarella from the front, the other from the back. Arabella doesn't realize this. She thinks twice "She is a spy." It is intuitively clear that her two thoughts are different even though they share character and express the same proposition. Searle develops an account of indexicals that distinguishes the two thoughts. His view is that the content of Arabella's thought is something like [the (female) person who is causing *this* visual experience is a spy].<sup>18</sup> One way of understanding Searle's view is that the proposition expressed by this thought contains the visual experience as a constituent. Searle's semantics is apparently solipsistic since the interpretation of the thought requires the existence of nothing other than Arabella and her mental contents.

Hilary Putnam first introduced twin arguments (like the one used above) to show that natural kind terms, expressions like "water," "gold," "tiger," and so forth, have nonsolipsistic

meanings.<sup>19</sup> Arabella and Twin Arabella are neurophysiologically identical. They inhabit, respectively, Earth and Twin Earth, which are identical except that the stuff called "water" on Earth is composed of H<sub>2</sub>O molecules, whereas the stuff called "water" on Twin Earth is composed of XYZ molecules. We also suppose that the time is before chemistry has been discovered so that no one on Earth or Twin Earth can distinguish H<sub>2</sub>O from XYZ. According to Putnam, Arabella's word "water" refers to H<sub>2</sub>O, whereas Twin Arabella's word "water" refers to XYZ. A version of the causal theory of reference explains these reference relations. Roughly, Arabella's use of "water" refers to H<sub>2</sub>O because her use is a link in a causal chain that begins with original dubbings of samples of H<sub>2</sub>O. Arabella's use refers to anything that belongs to the same natural kind as the original samples. Since XYZ is not the same natural kind as H<sub>2</sub>O, Arabella's tokens of "water" do not refer to XYZ, even though Arabella cannot distinguish the two kinds. If she were miraculously transported to the shores of the Twin Pacific on Twin Earth and said "Water, water, everywhere," she would be wrong.

When Arabella and Twin Arabella think the thoughts each would express by uttering "Water is wet," they think different thoughts. Arabella's thought is true iff H<sub>2</sub>O is wet, whereas her twin's thought is true iff XYZ is wet. It follows that their thought contents do not supervene on their neurophysiologies. Furthermore, on Putnam's causal account of how "water" gets its meaning, Arabella could not think her thought unless she was on the receiving end of a causal

chain that originates with an event involving H<sub>2</sub>O. Putnam's interpretation of and theory of meaning for natural kind terms are squarely nonsolipsistic.<sup>20</sup>

Putnam's positive account is that the meaning of a natural kind term consists of a number of components. One is the reference of the term, in our example, H<sub>2</sub>O. A second component he calls "stereotype." It consists of the information that competent speakers associate with water, for example, that it quenches thirst, fills oceans, and so forth. Putnam gives the impression that this component is entirely within the mind. This suggests the possibility that Arabella's mental representation corresponding to "water is wet" may have two interpretations: a nonsolipsistic interpretation that includes H<sub>2</sub>O, and a Fregean solipsistic interpretation as in [the stuff that quenches thirst, fills oceans, and so forth, is wet]. We will pursue this idea later.

Tyler Burge pushes a variant of the Twin Earth parables that also shows that the meanings of certain expressions are determined by factors external to the thinker.<sup>21</sup> Burge imagines an English speaker who does not know that arthritis is specifically a condition of the joints, although most of her beliefs concerning arthritis are true. She utters "I have arthritis in my thigh." According to Burge, her utterance means that she has arthritis (in our sense) in her thigh. Her reason for claiming this is that Arabella will defer to members of her linguistic community should they correct her. When corrected, she will say that her utterance was false. Burge then considers this woman's twin, who speaks Twin English, which is like English, except

that in it, "arthritis" refers to inflammations of the thigh as well as of the joints. Her twin's utterance is true. On the assumption that the thought expressed by an utterance has the same interpretation as the utterance, Arabella and her twin think different thoughts even though they are neurophysiologically identical. [This we take to be Putnam's early view.]

It is interesting to compare Burge's and Putnam's arguments. Putnam's argument applies to natural kind terms. Burge's argument can apparently be applied to any expression, even to adjectives, adverbs, and logical connectives. If Putnam is correct, then the interpretation of a natural kind term includes the reference of the term - for example, the substance water. If water never existed, we couldn't think that water is wet anymore than we could think that Aristotle is wise if Aristotle never existed. Burge's argument has no such conclusion. The term "arthritis" might have been introduced by description and there might never have been cases of the disease.<sup>22</sup> Putnam and Burge have nonsolipsistic theories of meaning, but they emphasize different ways in which meaning is determined. Putnam emphasizes the causal connections between the tokening of an expression and a dubbing of a natural kind. Burge emphasizes the role that one's linguistic community has in determining the meanings of one's words. Of course, it may be that both Putnam and Burge are correct and that an adequate theory of meaning for English will include reference to both causal chains and community practices.

The arguments that we have quickly canvassed purport to show that certain expressions and representations have nonsolipsistic semantics. In each case, the arguments show that these representations have meanings that they could not have unless certain individuals other than a thinker and her thoughts exist. How might the view that thought possesses *SS* be defended against these arguments? One strategy is to fight battles on each front, arguing that Kripke is mistaken about names, Putnam mistaken about natural kind terms, and so forth. This is the strategy pursued by Searle. The other strategy is to admit defeat on the fronts but then to circle the wagons around the mind and defend the possibility of constructing a characterization of meaning that is solipsistic. This is the strategy pursued by Fodor and the one that we will follow. Still, we can make use of Searle's accounts by taking his views about the correct semantics of English expressions as suggestions for how to construct *SS*

Fodor's strategy is to associate with each mental representation a narrow and a wide content.<sup>23</sup> The arguments of Kripke, Putnam, et. al., are taken to show that wide content does not supervene on the thinker's body. But narrow content is supposed to supervene on the thinker's body, including states of his brain and sense organs, and therefore is a version of physicalistic *SS*. The problem is to construct an appropriate notion of narrow content.

When he wrote *MS*, Fodor seemed to think that it would not be all that difficult to construct *SS* for mental representations that could play an explanatory role in cognitive

psychology. After discussing some of the antisolipsistic considerations we have reviewed, he remarks:

To summarize: transparent taxonomy is patently incompatible with the formality condition: whereas taxonomy in respect of content may be compatible with the formality condition, plus or minus a bit. That taxonomy in respect of content is compatible with the formality condition, plus or minus a bit, is perhaps *the* basic idea of modern cognitive theory [emphasis in the original].<sup>24</sup>

Fodor is claiming that if we stick to opaque, as opposed to transparent, interpretations of mental representations, we will come close to semantics that conforms to the formality condition, that is, to *SS*. But it seems to us that the construction of *SS* that can be used in cognitive theory is a much more formidable, perhaps an impossible, task. In the remainder of this paper we will consider reasons why this is so.

What are the adequacy conditions that a characterization of narrow content must satisfy? Since it is solipsistic, it will assign the same contents to Arabella's and her twin's thoughts "Water is wet" and also assign the same contents to the thoughts of the woman and her twin in Burge's story. It will assign different contents to the thoughts expressed by "Cicero was bald" and "Tully was bald" when the thinker does not believe that Cicero = Tully. The characterization

of narrow content should serve the needs of cognitive theory. Fodor seems to understand this requirement so that propositional attitudes interpreted narrowly will yield rationalizing explanations of action (when the actions themselves are described narrowly). Folk psychological theory, according to Fodor, contains generalizations like the following: when an individual believes that his obtaining water requires that he raise his hand and he wants it to be the case that he obtains water, then he will, *ceteris paribus*, raise his hand. This generalization does not apply to Twin Earthlings because they do not have beliefs about water. This is due to the fact that belief content is characterized widely in the generalization. Fodor suggests that cognitive theory will contain refinements of such generalizations which apply on Earth and Twin Earth. So a requirement on narrow content ascriptions is that it should be employable in such generalizations. In MS, Fodor held that the appropriate notion of content satisfies the formality condition and so supervenes on neurophysiological states. In more recent writings, he seems to hold that narrow content supervenes on bodily states or perhaps on bodily states together with a specification of inputs to the organism's perceptual systems.<sup>25</sup> If the inputs are characterized in ways that make reference to no individuals external to the organism's body, then the characterization is still a version of SS. One final requirement on narrow content is for there to be a plausible theory of meaning that is solipsistic and that accounts for how representations obtain their contents.

What form will *SS* take? One currently fashionable answer is provided by "conceptual role theories" (*CRT*).<sup>26</sup> A *CRT* for a person's language of thought characterizes the meaning of a mental representation in terms of its causal or inferential role in relating stimuli, behavior, and the tokening of other mental representations. The characterization is a version of *SS* only if representations, behavior, and stimuli are described in ways that make no reference to individuals external to the thinker's body. We have discussed *CRTs* elsewhere and argued that characterization of *CR* is not itself a characterization of content.<sup>27</sup> For present purposes it is sufficient to point out that a characterization of *CR* does not yield appropriate complements to put into "believes that . . ." and other propositional attitude contexts. But we need such expressions of content if we are to construct rationalizing explanations of behavior of the sort that Fodor wants to capture in cognitive psychology. It should also be clear that *CR* does not provide a characterization of content suitable for expressing Cartesian skepticism. Descartes was not claiming that all his thoughts might be false even if they have the same conceptual role they actually have. They could have the same conceptual role and yet be about quite different things. So we have to look elsewhere for a specification of narrow content.

We know of two other proposals for constructing narrow content or *SS* for linguistic and mental representations. One we will call the "indexicalist strategy." It involves interpreting thoughts indexically in a way that is supposed to presuppose no individual external to the

thinker. We will call the second approach the "phenomenological strategy." It involves finding a collection of expressions that have *SS* and constructing interpretations for thoughts from these expressions. The approach is called "phenomenological" since the nonlogical vocabulary of these interpretations consists of "observation" terms that are supposed to describe how things seem or how they appear. In a number of recent papers, Fodor has employed both proposals to construct a characterization of narrow content. But we will argue that the prospects for success are poor. We will discuss the phenomenological strategy first.

The phenomenological strategy applied to proper names suggests that a name is interpreted as expressing a sense that the thinker associates with the name. For example, the thought that Aristotle was Greek might be interpreted as having the content that the author of the *Metaphysics* was Greek (or some similar content that can be expressed without using the proper name "Aristotle"). The phenomenological strategy might be applied in the following way to natural kind terms. Although Arabella and Twin Arabella refer to different things when they utter "Water is wet," it may be that they associate the same stereotype with "water." This suggests that the stereotype has a solipsistic interpretation. The idea is that the thoughts of both twins can be interpreted phenomenologically as having a content like [the liquid that people drink, fills oceans, and so forth, is wet].

However, the interpretations that we associated with "Aristotle was Greek" and "Water is wet" are certainly not completely solipsistic. The description "The author of the *Metaphysics*" contains another proper name and so we do not yet have a solipsistic interpretation. And even if the name were replaced by a description, the question would arise whether the predicates that occur in the description can be given solipsistic interpretations. This question also arises when we consider the stereotypes associated with kind terms. The suggestion was that the solipsistic interpretation of "Water is wet" is that the liquid that people drink, fills oceans, and so forth, is wet. It is clear that the content of this stereotype is not sufficiently narrow to be solipsistic. The expressions "oceans" and "people" have different meanings for Arabella and Twin Arbella. For Arabella, "people" refers to Earthlings, whereas for Twin Arabella it refers to Twin Earthlings. The same point applies to "oceans" and "liquid" and perhaps to other concepts in the stereotype. By imagining suitable differences between Earth and Twin Earth, while keeping constant the ways things seem to the twins, it looks as though Twin Earth arguments will succeed in showing that no natural kind term has *SS*. It might be suggested that those expressions that describe the ways things seem, the truly phenomenological expressions, are immune from the Twin Earth arguments. Are there any such predicates in English? The best candidates are "observation terms," for example, "red," "round," and "bitter." If these do not have *SS*, it is difficult to see how the phenomenological can be made to work.

As we already pointed out, the argument Burge gave to show that the meaning of "arthritis" depends on features external to the thinker, specifically, community usage, applies to any natural language expression. If these arguments are correct, they show that even predicates like "is red" do not have SS. However, it seems that we can imagine a language that is like English except that the deferential practices on which Burge's arguments rely are absent. For this reason we will give another argument that shows that the semantics of observation terms is nonsolipsistic. Suppose that on Twin Earth those things that are red on earth (blood, ripe tomatoes, boiled lobsters - or rather their counterparts on Twin Earth) are green. So, if an earthling visited Twin Earth, she would correctly think that the things that Twin Earthlings call "boiled lobsters" are green. However, the Twin Earthlings are born with color-inverting lenses so that when looking at what they call "a lobster," they experience the same kind of sensations (are in the same brain states) that Earthlings experience when looking at boiled lobsters. Suppose, as usual, that Arabella and her twin are neurophysiologically type identical and that each utters "Roses are red." The things the Twin Earthlings call "roses" are actually green.

How should we translate Twin Arabella's word "red" into English? Her utterances of "That's a red one" are typically caused by things that are green. When she says "I am looking for a red dress," she is satisfied when she finds a green one. So we have every reason to suppose that her word "red" means green. If we translate her word "red" by our word "red," the result

would be that we would interpret Twin Arabella as being pervasively mistaken about the colors of things. This is certainly intolerable. It is much more plausible to translate her "red" by our "green." This translation interprets that Arabella and Twin Arabella are thinking different thoughts when each says to herself "That's a red one," even though they are in neurophysiologically identical states and they are experiencing the same qualia.<sup>28</sup>

Fodor might reply that Arabella and her twin are really in identical bodily states since the twin's color-inverting glasses count as part of her body. In his most recent discussions of narrow content, Fodor characterizes narrow content so that it supervenes on the states of an organism's brain and transducers.<sup>29</sup> Plausibly, the color-inverting lenses are part of the twin's visual system. This reply can be deflected with some more science fiction. We suppose that on Twin Earth a substance in the atmosphere changes light from red to green and vice versa soon after it is reflected. In the revised story, Arabella and her twin are in identical brain and transducer states although one is thinking that's a red one while the other is thinking that's a green one.<sup>30</sup>

The second approach that the solipsistic semanticist can take is the indexicalist strategy. An indexical sentence has both a content and a character. The character is a function from contexts to contents. So when Arabella and Twin Arabella each utter "She is a spy," their utterances have different contents (one utterance is about Barbarella, the other about twin Barbarella) but the same character. Character may be solipsistic even if content is not. So if we

can associate an indexical interpretation with each mental representation, we might yet succeed in constructing *SS* for thought. But we doubt that this can be carried out. Exactly what indexical interpretation can be given to "Water is wet"? Fodor suggests that "perhaps 'water' means something like 'the local, transparent, potable, dolphin-torn, gong tormented . . . stuff one sails on.'"<sup>31</sup> The reference to its being local provides the indexicality. Arabella and her twin may mean the same by "water" and yet refer to different substances because they inhabit different contexts. But the obvious problem with this suggestion is that this paraphrase contains expressions that are interpreted nonsolipsistically.

Searle suggests an interpretation for "water" that may seem to avoid this problem. He says that "'water' is defined indexically as whatever is identical in structure with the stuff causing *this* visual experience."<sup>32</sup> But this suggestion faces a couple of difficulties. First, it is not clear that "stuff" and "causing" or even "visual experience" have nonsolipsistic semantics. Second, the expression does not uniquely refer. There are many things and events causing *this visual experience*. One of the causes is the pattern of neuron firing in the optic nerves. We could exclude this by adding "the external liquid stuff causing. . .," but now we are faced with the problem that "external liquid" does not have *SS*.

There is another way to pursue the indexicalist strategy.<sup>33</sup> Think of Arabella's entire environment, including its history, as a context. Her sentence "Water is wet" can be interpreted

as expressing a character that maps that context onto the content that H<sub>2</sub>O is wet. Twin Arabella's sentence expresses the same character, which maps her different context onto the content that XYZ is wet. It is clear that character construed in this way supervenes on neurophysiology. But it cannot serve as a specification of narrow content. At most, the account provides sufficient conditions for when two individuals' thoughts have the same meaning, and then, only when the two are neurophysiologically identical. If two people are not neurophysiologically identical, then the account says nothing concerning whether structurally similar representations possess the same or different characters. Nor does this characterization of meaning as character yield appropriate specifications of content that can follow "believes that." As Fodor, who suggests this proposal, says, "First it is one thing to have a criterion for the intentional identity of thoughts; it is quite another to be able to say what the intention of a thought is."<sup>34</sup> It is the latter that we need if we want to employ narrow content in the rationalizing explanations of cognitive psychology.

The preceding considerations show that it is not easy to construct *SS*. The usual, nonsolipsistic interpretation of thoughts misses being solipsistic by a great deal more than "plus or minus a bit." There is no solipsistic part of English out of which solipsistic interpretations of thought can be constructed. But this doesn't show that *SS* is impossible. It shows only that our language is so thoroughly nonsolipsistic that it doesn't contain the resources to construct nonsolipsistic semantics. Perhaps mental representations have *SS* that cannot be expressed in

English. However, we will argue that the most plausible theories of meaning for mental representations rule out *SS*. Since Fodor himself advocates a version of this theory of meaning, our argument will be, to a certain degree, an *ad hominem* one.<sup>35</sup>

The account that Fodor favors is a development of Fred Dretske's views. According to Dretske, the content of a mental representation is determined by its informational origins.<sup>36</sup> To consider the simplest kind of case, suppose that a mental structure can either be in state *Y* or state *N* and that under certain "normal conditions," it is in *Y* when the organism is looking at a fly but in *N* when the organism is not looking at a fly. Then the mental structure carries the information that there is a fly in view when conditions are normal. Fodor endorses an account of this kind although he substitutes certain epistemically ideal conditions for normal conditions.<sup>37</sup>

Although there are substantial difficulties with Dretske's and Fodor's proposals, it is the most promising.<sup>38</sup> Here we only want to show that if we take this account seriously, then either it results in nonsolipsistic semantics or it leads to *SS* that interprets one's mental states as being about one's own nervous system. First, observe that on a plausible understanding of "normal" and "ideal," Fodor's theory of meaning is already nonsolipsistic. Normal and ideal conditions are such in virtue of the evolutionary history of an organism. If the organism had evolved in different circumstances, different conditions might count as normal or ideal. Let's ignore this problem for a moment. Suppose that a frog's mental state carries the information that a fly is in

view. Clearly it carries this information only because of its interactions with its environment. Change the environment, and the frog's neural state will carry different information, for example, that a moving BB is in view. The obvious reply to this point is to say that the frog's neural state was designed by evolutionary pressures to carry information about the presence of flies not BBs. But the proponent of *SS* cannot avail himself of such considerations because the self-assigned task is to construct meaning from ingredients that are entirely within the head. Neurophysiologically identical organisms might have evolved in different environments and so their neural states might carry different information. If we are to use a Dretskean theory of meaning to support a solipsistic interpretation of the frog's mental representations, then we will have to find something for these representations to carry information about that is immune from the kind of environmental tinkering used in the Twin Earth stories. But it would seem that all that can remain constant under these environmental changes is the working of the frog's nervous system itself. There will be some covariation between states *Y* and *N* and certain patterns of irradiation on the frog's eyes. So we can solipsistically interpret *Y* as having the content that a certain pattern of ocular irradiation is occurring. If we adopt a Dretskean theory of meaning, then the only interpretations consistent with *SS* interpret an organism's thoughts as being about its own nervous system.

A proponent of *SS* holds that, although thoughts and ideas purport to refer to the external world, their meanings are entirely a product of mental activity. Originally, he may have thought that the semantics for all the expressions of our language is solipsistic. But the considerations advanced by Kripke, Putnam, and Kaplan show that vast portions of natural language have interpretations that are incompatible with *SS*. This showed that our usual ways of individuating the contents of mental representations are also nonsolipsistic. We considered a response that granted the points made by Kripke, et al., but attempted to construct solipsistic interpretations out of the fragment of language that remained solipsistic. We argued that this strategy is unlikely to succeed because almost all natural language expressions, even observation terms, have nonsolipsistic semantics. The prospects for constructing *SS* for thought seem even bleaker when we reflect on the difficulty of providing a plausible account about what makes it the case that a thought has the content it has that is compatible with solipsism. If an informational answer to this question of the sort suggested by Dretske and Fodor is correct, then a language with *SS* is a language that is solipsistic in another way. In it one can refer only to oneself and one's mental states. Solipsism with respect to sense results in solipsism with respect to reference.<sup>39</sup>

If the view that thought possesses *SS* is not plausible, what are the consequences for cognitive theory and for our Cartesian intuitions? With respect to cognitive theory, there seem to be two alternatives. The first, recommended by Stich, is to abandon the use of propositional

attitude explanations in cognitive theory.<sup>40</sup> The second, recommended by Burge, is to accept the fact that propositional attitudes do not supervene on bodily states and to argue that this in no way counts against their explanatory ability.<sup>41</sup> We cannot enter into this debate here. But we do want to make a few remarks concerning the Cartesian intuitions that seem to provide such strong support for there being *SS* for thought.

The Cartesian observes that it is possible for all his thoughts about the external physical world to have the contents they have even if there is no external world. If this is correct, then thought-world interactions cannot be an essential determinant of the contents of thoughts. It follows that thought has *SS*. The first point to make about this argument is that at most it shows that it is *epistemically* possible for the Cartesian that he has the thoughts he actually has even if there is no external world. The argument is similar to other Cartesian arguments that show that it is epistemically possible for a mental event or state to exist even though no physical events or states exist. But epistemic possibility is not the same as metaphysical or conceptual possibility and only these would establish the need for *SS*. However, we must admit that our arguments do not establish that *SS* is not possible. At most we have shown that plausible theories of meaning are nonsolipsistic.

There is a related intuition that may lead the Cartesian into thinking that thought must have *SS*. It is the observation that one can know the contents of one's thoughts without engaging

in any empirical investigation. When considering another person, we might be persuaded that his thoughts have the contents they have in virtue of interactions with his environment, his linguistic community, and so forth. But I can know the contents of my own thoughts without evidence and, in particular, without knowing much about these matters. Since I can know the contents of my thoughts without investigating the world outside myself, it is tempting to conclude that these contents must be determined by events entirely within myself. We have here a real and, we think, quite deep tension. The tension between nonsolipsistic theories of meaning and the Cartesian intuitions would be relieved somewhat if it could be shown that there is no genuine conflict between the claim that we know our own thought contents without evidence and the claim that those thought contents are determined by matters external to us about which we may have no knowledge.

We would like to show that even though a mental representation  $R$  has the content it has in virtue of matters external to the thinker, he might still know  $R$ 's content without evidence. We tentatively offer the following account. Let us suppose that  $A$  grasps the concept *water*. We cannot give necessary and sufficient conditions for grasping this concept but something like the following story seems plausible. To grasp the concept *water* is to have a mental representation that plays a certain kind of functional role in one's thought and that is related in an appropriate way to  $H_2O$ .<sup>42</sup> Suppose that  $A$  thinks the thought that water is wet. What is required of  $A$  in order

that he know what his thought means? If it is required that he knows that water is H<sub>2</sub>O, then it must be admitted that he does not know what his thought means. But this requirement is not plausible and does not follow from the fact that *A* grasps the concept. Now suppose that *A* grasps the concept is *true* and has a way of referring to his thoughts, say by quoting them. *A*'s thought ["Water is wet" is true iff water is wet] will then express his knowledge of the content of his thought "Water is wet." And *A* will know that this thought is true simply in virtue of grasping the concepts of *truth*, *quotation*, *water*, and *is wet*. Suppose that *A* had lived on Twin Earth. In that case, his concept would be different since it would refer to XYZ. However, he could know the content of his thought "Water is wet" since he knows the truth conditions. He would express this knowledge by saying " 'Water is wet' is true iff water is wet."

It might be objected that in our example *A* does not know that "Water is wet" is true iff water is wet but only that the representation " 'Water is wet' is true iff water is wet" is true.<sup>43</sup> The source of the objection is this. Suppose that *B* knows how quotation works in English, knows the disquotational effect of "is true," and is able to recognize grammatical sentences of English but knows nothing else about English. *B* will be in a position to recognize English sentences of the form " 'Water is wet' is true iff water is wet" as true. But this is not the same as knowing the truth conditions of "Water is wet."<sup>44</sup> Someone could recognize this sentence as true and yet not know what it means. The objection is that *A* is in the same position with regard to his

thoughts as *B* is with respect to English. Our reply is that there is an important difference between the two cases. We assumed that *A* had the ability to think that water is wet on the basis of there being a mental representation that plays a certain conceptual role in his thinking and its playing this role in a particular environment. The sentence "Water is wet" does not play the appropriate functional role in *B*'s thought that it would if *B* understood "Water is wet." But *A*'s thought "Water is wet" does play the appropriate role; it must if it is to be the thought that water is wet. If this is correct, then we can see how the meaning of a person's thoughts can depend on matters external to him and yet how he can be in a position to know the meaning of a thought without appealing to these external matters <sup>45</sup>

#### Notes

1. René Descartes, *The Philosophical Works of Descartes, vol. 1*, translated and edited by Elizabeth S. Haldane and G. R. T. Ross (Cambridge, 1967), 146.
2. Some qualification is needed here. Descartes did think that some of our ideas were not conceptually independent of anything outside the mind, for example, our idea of God and of infinity.

3. The view that ideas are images and that images represent by picturing was held by Locke. See his *Essay Concerning Human Understanding*, bks. 2 and 3.
4. David Hume, *A Treatise on Human Nature*, edited by Selby-Bigge (Oxford, 1888), 20.
5. Gottlob Frege, *Posthumous Writings*, edited by H. Hermes (Berkeley, 1979), 191.
6. A property P supervenes on properties Q 1, Q2, . . . , if and only if it is metaphysically impossible for two individuals to differ with respect to P without differing with respect to some of the Qs. For a discussion of supervenience, see Jagwon Kim, "Concepts of Supervenience," *Philosophy and Phenomenological Research* 45 (1984): 153-76.
7. John Searle, "Minds, Brains, and Programs," *The Behavioral and Brain Sciences* 3 (1980), and *Intentionality* (Cambridge, 1983). Hereafter these works are indicated by BBS and INT.  
Jerry Fodor, "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," *Representations* (Cambridge, Mass., 1980), 225-56. Hereafter MS.
8. MS 229.
9. MS 250.
10. René Descartes, *Descartes' Philosophical Writings*, translated and edited by E. Anscombe and P. Geach (New York, 1971), 78.

11. Saul A. Kripke, *Naming and Necessity* (Cambridge, Mass., 1980). Kripke does not commit himself to the view that there is a causal relation between the token of a name and its bearer, although others have made this claim.
12. Saul Kripke, "A Puzzle about Belief," *Meaning and Use*, edited by A. Margalit (Dordrecht, 1976), 239-83.
13. Kripke can avoid this objection by abandoning transparency.
14. Kripke's nonSS requires either that we cannot have such thoughts or that the mental representation corresponding to "Homer" is not a directly referring expression.
15. David Kaplan, "Demonstratives," unpublished manuscript, 1977.
16. One might infer from this nonsupervenience that Kaplan holds a nonSS theory of meaning. However, compare note 18. Supervenience fails and yet SS obtains. Nothing in Kaplan's writings tells whether he is or is not a nonSS theorist of meaning about indexicals. We also note that transparency fails for him. Arabella may be looking in a mirror, pointing unwittingly at herself, and think what she would express by "I am tall" and "She is tall," not knowing these express the same proposition.
17. Actually, we think something much stronger than this. In our "Dual Aspect Semantics," forthcoming, we argue that, at best, character provides a method for individuating thoughts, not for identifying or ascribing them.

18. John Searle, INT: 203. Interestingly, on Searle's account Arabella and her twin are having thoughts with different contents since each refers to her own visual experience. So even though Arabella and Twin Arabella are in type identical neurophysiological states, they can have solipsistic thoughts with different contents. This shows that we cannot always argue from the fact that two individuals in type identical neurophysiological states have thoughts with different contents to the conclusion that the thoughts are nonsolipsistic.
19. Hilary Putnam, "The Meaning of 'Meaning'," in *Mind, Language, and Reality*, vol. 2 (Cambridge, 1975) 215-71.
20. Putnam also abandons transparency. The thoughts expressed by "Water is wet" and "H<sub>2</sub>O is wet" are the same, but surely one need not recognize them as such.
21. Tyler Burge, "Individualism and the Mental," *Midwest Studies* 4 (1979): 73-121.
22. This suggests we could have a language with a nonSS theory of meaning but with SS interpretations.
23. Jerry Fodor, "Narrow Content and Meaning Holism," unpublished manuscript.
24. Jerry Fodor, MS 240.
25. Jerry Fodor, "Narrow Content and Meaning Holism."
26. See, for example, Brian Loar, *Mind and Meaning*, (Cambridge, 1981), Colin McGinn, "The Structure of Content," *Thought and Object*, edited by A. Woodfield (Oxford, 1982), 215-71,

and Hartry Field, "Logic, Meaning, and Conceptual Role," *Journal of Philosophy* 7 (1977): 379-409.

27. See Ernest LePore and Barry Loewer, "Dual Aspect Semantics," forthcoming.

28. We do not mean just that we lack epistemic warrant in translating her word "red" as our word "red" but that it would be an error to do so.

29. Jerry Fodor, "Narrow Content and Meaning Holism."

30. Fodor's strategy suggests a fourth semanticist. One might be a *SS* theorist of meaning but have non*SS* interpretations. This depends on several things. First, it is important to see that Fodor is not offering a theory of meaning here. The phenomenological approach is a strategy for ascertaining which thoughts an individual can entertain without looking outside the organism's physical structure. If this strategy succeeded, one might be inclined to infer that there is a *SS* theory of meaning for mental representations. However, Fodor has not given us one. He is not here telling us *in virtue of* what do mental representations have the contents they do. He is telling us only how to discern these contents. Second, Fodor is not a phenomenalist. He does not claim that phenomenological properties are in the head. This is no problem for him. *Qua* cognitive psychologist, his concern is only that he need not look outside the organism to discern its psychology. However, that these properties exist outside the organism does not commit Fodor to non*SS* interpretations. This depends on his ontological view about

properties. If properties are universals, abstract entities, then his interpretations are *SS*. If phenomenological properties are sets of physical objects, then his interpretations presuppose the existence of physical entities outside the organism, and he is therefore committed to non*SS* interpretations.

31. Jerry Fodor, "Banish Discontent," unpublished manuscript.

32. John Searle, INT: 203.

33. A view like this one is developed by Stephen White in "Partial Character and the Language of Thought," *Pacific Philosophical Quarterly* 63 (1982): 37-65.

34. Jerry Fodor, "Banish Discontent."

35. We are not claiming that Fodor ever intended Dretske's informational theory of meaning to be a *solipsistic* theory of meaning of these representations. However, this is the only kind of theory of meaning that Fodor has presented for mental representations.

36. Fred Dretske, *Knowledge and the Flow of Information* (Cambridge, Mass., 1981), chaps. 10 and 11.

37. Jerry Fodor, "Psychosemantics," unpublished manuscript.

38. Informational accounts of content are discussed by Barry Loewer in "Information and Content," *Synthese*, forthcoming.

39. This is not necessarily a criticism. Someone might simply swallow the idea that all our thoughts are about our own mental states in order to keep meaning in the head. But Fodor cannot if he is to provide rationalizing explanations of the sort discussed above. No matter how narrowly we describe Arabella's and Twin Arabella's behavior, we will not be able to construct a cognitive psychology that issues in rationalizations of their behavior if we limit these twins to having thoughts only about their own brain states.
40. See Steven Stich, *From Folk Psychology to Cognitive Science* (Cambridge, Mass., 1983).
41. Tyler Burge, "Individualism and Psychology," read at Sloan Conference, M.I.T., 1984.
42. This kind of account of concepts is developed by Colin McGinn in "The Structure of Content," *Thought and Object*, edited by A. Woodfield (Oxford, 1981).
43. This objection was made in conversation by Paul Boghossian, who is unpersuaded by our reply.
44. Much is made of this distinction in Barry Loewer and Ernest LePore, "Translational Semantics," *Synthese*, 1980.
45. We would like to thank Paul Boghossian, Donald Davidson, Umberto Eco, Peter Klein, Hilary Putnam, Steven Shiffer, and Bas van Fraassen for discussion on earlier drafts of this paper.

Earlier versions of this paper were read at the Universities of Alberta, Calgary, Oklahoma, Michigan, Central Michigan, Regina, and the Florence Center for the History and Philosophy of Science. We would like to thank all the members of these various philosophy departments for their helpful suggestions and criticisms.