Mind

A Brief Introduction

John R. Searle

OXFORD
UNIVERSITY PRESS
2004

CHAPTER SIX

Intentionality

The problem of intentionality is second only to the problem of consciousness as a supposedly difficult, perhaps impossibly difficult, problem in the philosophy of mind. Indeed, the problem of intentionality is something of a mirror image of the problem of consciousness. Just as it is supposed to be extremely difficult to fathom how mere bits of matter inside the skull could be conscious, or could through their interactions create consciousness, so it is difficult to imagine how mere bits of matter inside the skull could "refer to" or be about something in the world beyond themselves, or could through their interactions create such a reference. To take an example, I am now thinking that the sun is 93 million miles from the Earth. My thoughts definitely refer to, or are about, the sun. They are not about the moon, my car in the garage, my dog Gilbert, or my next-door neighbor. Now what is it about the thought that enables it to reach as far as the sun? Do I send mental rays all the way to the sun, just as the sun sends light rays all the way to the Earth? Unless there is some sort of

connection between me and the sun, it is hard to imagine how my thoughts could reach the sun. And what goes for the sun, goes for just about any object that I can represent in my beliefs, desires, and other intentional states. So for example, if I think that Caesar crossed the Rubicon, then my thought is about Caesar, and it has the content that he crossed the Rubicon. But now, what fact about the stuff inside my skull makes it refer all the way back in history to a specific individual, and a specific river, and ascribe the specific action of the individual crossing the river?

In addition to the problem of how such a thing is possible, there is a related problem about how I can be so confident that it is happening right. When I refer to Julius Caesar how can I be so smugly confident that my thoughts are actually hitting Julius Caesar and not, for example, Mark Anthony or Caesar Augustus or my dog Gilbert? If I throw a stone into the dark, I may not have the faintest idea what it is hitting, but when I throw my reference into the unseen I am often completely confident about what it is hitting.

To make matters even worse, it seems that I can sometimes think about objects that do not even exist. When I was a small child I believed that Santa Claus comes on Christmas Eve. Was my belief about Santa Claus? It certainly seems so, but how is that possible, since Santa Claus does not even exist?

Notice that these are questions that only a philosopher would ask. Philosophy begins with a sense of mystery and wonder at what any sane person regards as too obvious to worry about.

Notice also that we cannot explain the intentionality of the mind by saying it is just like the intentionality of language. In the case of language, the utterance "Caesar

crossed the Rubicon" is about Caesar and says of him that he crossed the Rubicon. I cannot say that a mental representation derives its intentional capacity from language, because of course the same problem arises for language. How is it possible that a mere sentence, sounds that come out of my mouth or marks that I write on paper, can refer to, be about, or describe objects and states of affairs that are 2,000 years in the past and 10,000 miles away? The intentionality of language has to be explained in terms of the intentionality of the mind and not conversely. For it is only in virtue of the fact that the mind has imposed intentionality on these sounds and marks that they refer to the objects and events that I have mentioned. The meaning of language is derived intentionality and it has to be derived from the original intentionality of the mind.

There are three problems about intentionality we need to address. First, how is intentionality possible at all; second, given that intentional states are possible, how is their content determined; and third, how does the whole system of intentionality work? Most of the philosophical literature is about the first two questions. I find the third question the most interesting. In this chapter I am going to first deal with the question about how intentionality is possible, and I will use my usual method of trying to demystify the whole phenomenon by bringing it down to earth. Then I will go to the third topic and describe the structure of intentionality; and I will include a section on the differences between intentionality-with-a-t and intensionality-with-an-s. Finally, I will conclude with the second question, how the contents of intentional states are determined. Readers familiar with cognitive science will recognize that when we talk about intentionality we are

discussing what in cognitive science is known as "information." I prefer "intentionality" because "information" is systematically ambiguous between a genuinely observer-independent mental sense (for example, by looking out the window now I have information about the weather) and a nonmental observer-relative sense (for example, the rings in the tree stump contain information about the age of the tree). This ambiguity can also arise for "intentionality," but it is easier to avoid and confusion is less likely.

I. HOW IS INTENTIONALITY POSSIBLE AT ALL?

This problem is supposed to be as difficult as the problem of consciousness, so the sorts of solutions that are supposed to solve it are much like the solutions proposed for the problem of consciousness.

The dualistic solution is to say that as there are two different realms, the mental and the physical, so the mental realm has its own sorts of powers not possessed by the physical realm. The physical realm is incapable of referring, but the mental realm is essentially capable of thinking, and thinking involves reference. I hope it is obvious that this dualistic solution is no solution at all. To explain the mystery of intentionality it appeals to the mystery of the mental in general.

I think that the most common contemporary philosophical solution to the problem of intentionality is some form of functionalism. The idea is that intentionality is to be analyzed entirely in terms of causal relations. These causal relations exist between the environment and the agent and between various events going on inside the agent. On this view there is nothing mysterious about intention-

ality. It is just a form of causation. The only special feature is that intentional relations exist between the agent's cerebral innards and the external world. And, by this time, I do not need to tell the reader that the most influential version of functionalism is computer functionalism, or Strong Artificial Intelligence.

Finally, there is the eliminativist view of intentionality: there really are no intentional states. The belief that there are such things is just a residue of a primitive folk psychology, one that a mature science of the brain will enable us to overcome. A variant of the eliminativist view is what we might call "interpretativism." The idea here is that attributions of intentionality are always forms of interpretation made by some outside observer. An extreme version of this view is Daniel Dennett's conception that we sometimes adopt the "intentional stance" and that we should not think of people as literally having beliefs and desires, but rather that this is a useful stance to adopt about them for the purpose of predicting their behavior.[1]

I will not spend much time criticizing these various accounts of intentionality because I have already criticized the general thrusts of these arguments in earlier chapters. What I want to do, as I did with the problem of consciousness, is bring the whole issue down to earth. If you ask, how is it possible that anything as ethereal and abstract as a thought process can reach out to the sun, to the moon, to Caesar, and to the Rubicon, it must seem like a very difficult problem. But if you pose the problem in a much simpler form, how can an animal be hungry or thirsty? How can an animal see anything or fear anything? Then it seems much easier to fathom. We are speaking, as we did about consciousness, of a certain set of biological capacities of the

mind. And it is best to start with the biological capacities that are primitive—for instance, hunger, thirst, the sex drive, perception, and intentional action. In the last chapter I described some of the neurobiological details about how brain processes cause conscious feelings of thirst. But in explaining how brain processes can cause feelings of thirst, we have already explained how brain processes can cause forms of intentionality, because thirst is an intentional phenomenon. To be thirsty is to have a desire to drink. When the angiotensin 2 gets inside the hypothalamus and triggers the neuronal activity that eventually results in the feeling of thirst it has *eo ipso* resulted in an intentional feeling. The basic forms of consciousness and intentionality are caused by the behavior of neurons and are realized in the brain system, that is itself composed of neurons. What goes for thirst goes for hunger and fear and perception and desire and all the rest.

Once we demystify the problem of intentionality by removing it from the abstract, spiritual level down to the concrete level of real animal biology, I do not believe that any unsolvable mystery remains about how it is possible for animals to have intentional states. If you start with such simple and obvious cases as hunger and thirst, intentionality is not at all difficult to explain. Of course, beliefs, desires, and sophisticated forms of thought processes are more complex and more removed from the immediate stimulation of the brain by the impact of the environment than are perceptions or feelings of hunger and thirst. But even they are caused by brain processes and realized in the brain system.

When it seems mysterious to us that intentional relations can exist at all, when we pose such questions as,

How is it possible for my thoughts to reach all the way to the sun or as far back in history as Julius Caesar? it is because we are imposing the wrong model of relations on the sentences that describe our intentional contents. Similarly, when we are puzzled about how we can have thoughts about things that do not exist at all, such as Santa Claus, it is because we are thinking of intentionality as if it were a relation like standing next to or hitting or sitting on top of. You cannot hit something that does not exist and you cannot sit on something that is 93 million miles away. But referring to or thinking about something is not at all like sitting on it or hitting it. It is rather a form of *representation* and the notion of representation does not require that the thing represented actually exist or that it exist in some immediate proximity to the representation of it. We ought to hear the question, How is it possible to think about Santa Claus if Santa Claus does not even exist? as like the question, How is it possible to make up a story about Santa Claus if Santa Claus does not even exist? There we have an easier problem because we see that it does not seem metaphysically difficult to make up fictional stories. When I say this I am not, of course, solving the problem because, strictly speaking, the intentionality of the story derives from the intentionality of the mental content. I am trying to remove a sense of mystery by showing how the apparently mysterious is like the obviously unmysterious. Our ability to have intentional contents about the nonexistent seems mysterious, but our ability to construct fictional stories seems much less mysterious.

However, there are a lot of other problems. For example, what is the relation between conscious and unconscious intentionality and how do intentional states

get the content they have? I will have to work my way up to the point where I can answer these questions. At this point, it seems to me the best thing I can do is describe the formal structure of intentional states, because we will not get a grasp on how intentionality functions, until we see the structural features of intentional states, such as beliefs and desires, hopes and fears, perceptions, memories, and intentions.

II. THE STRUCTURE OF INTENTIONALITY

1. Propositional Content and Psychological Mode

Because intentional states are capable of referring to objects and states of affairs in the world beyond themselves, they must have some sort of *content* that determines this reference, and indeed we need to distinguish the content of the state from the type of state that it is. Thus I can believe that it will rain, hope that it will rain, fear that it will rain, or desire that it will rain. In each case there is the same content, that it will rain, but that content relates to the world in different psychological modes: belief, fear, hope, desire, etc. This distinction, by the way, exactly parallels the same distinction in language. Just as I can order you to leave the room, so I can predict that you will leave the room, and I can ask whether you will leave the room. In each case we have the same content, that you will leave the room, but it is presented in different sorts of speech acts. A good way to think of this is to think of the state as consisting in a psychological mode, such as belief or desire, with a propositional content, such as the proposition that it is raining. We can represent this as S(p), where the S stand for the mode or type of state and the

p for the propositional content. Such states are often called "propositional attitudes."

Not all intentional states have an entire proposition as their content. One might just admire Eisenhower or love Marilyn, and in such cases the intentional state just refers to an object. Such states can be represented as S(n), where the n names or refers to an object.

Notice that intentional representations are always under certain aspects and not others. For example, I might intentionally represent an object as the Evening Star and not as the Morning Star even though one and the same object is both. The aspect "celestial body that shines near the horizon in the evening" is not the same aspect as "celestial body that shines near the horizon in the morning." *Intentional states always have aspectual shapes,* because all representation is under aspects. This is an important point, because any theory of intentionality must account for aspectual shape, and some materialist theories are unable to do so. I mentioned in chapter 3 that functionalism was unable to distinguish between the desire for water and the desire for H_2O and this is because the causal relations on which functionalism relies to analyze intentionality do not have the aspectual shapes characteristic of genuine intentionality. We will see in chapter 9, on the unconscious, that any theory of the unconscious needs to account for the presence of aspectual shape when an intentional state is unconscious.

2. Direction of Fit

Intentional states, again like speech acts, are related to the world in different ways. It is the aim of a belief to be true,

and to the extent that the belief is true, it succeeds. To the extent that it is false, it fails. Desires, on the other hand, are not supposed to represent how the world is, but how we would like it to be. Thus, if I believe that it is raining, my belief will be true if and only if it is raining. But if I desire that it should rain, then my desire will be satisfied or fulfilled if and only if it rains. Though these look similar, there is a crucial distinction. In the case of the belief, the intentional state is supposed to represent how things are in the world. The belief is, so to speak *responsible for fitting the world.* But in the case of the desire, it is not the aim of the desire to represent how things are but rather how we would like them to be. In the case of the desire it is, so to speak, the *responsibility of the world to fit the content of the desire*. I am going to introduce a piece of jargon to describe this distinction. Where the mental state is responsible for fitting an independently existing reality, we can say that the mental state has the "*mind-to-world direction* of fit," or alternatively, it has the "*mind-to-world responsibility* of fit." The mental state fits or fails to fit how things really are in the world. Beliefs, convictions, hypotheses, etc., as well as perceptual experiences, all have this mind-to-world direction of fit. The most common expressions for appraising success in achieving the mind-to-world direction of fit are "true" and "false." Beliefs and convictions can be said to be true or false. Desires and intentions are not true or false the way beliefs are, because their aim is not to match an independently existing reality, but rather to get reality to match the content of the Intentional state. For that reason I will say they have the "*world-to-mind direction* of fit" or the "*world-to-mind responsibility* for fit."

Some intentional states, though they have a propositional content, do not have a direction of fit because it is not their aim either to match reality (the mind-to-world direction of fit) or to get reality to match them (the world-to-mind direction of fit). Rather, they take it for granted that the fit already exists. Thus, if I am sorry that I stepped on your foot, or I am glad that the sun is shining, I take it for granted that I stepped on your foot and that the sun is shining. About such cases, I say that the intentional states have the "*null* direction of fit." They "presuppose" a fitting relation rather than assert it or try to bring it about. I find it convenient to represent the mind-to-world direction of fit with a downward arrow thus: ↓; the world-to-mind fit with an upward arrow thus: ↑; and the null fit with the null sign thus: Ø

3. Conditions of Satisfaction

Whenever we have an intentional state that has a non-null direction of fit, the fit will either be achieved or not: the belief will be true, the desire will be fulfilled, the intention will be carried out or not, as the case might be. In such cases, we can say that the belief, desire, or intention is satisfied. What stands to the belief's being true is what stands to the desire's being fulfilled, is what stands to the intention's being carried out. I propose to describe this phenomenon by saying that every intentional state that has a non-null direction of fit has *conditions of satisfaction*. We can think of mental states as representations of their conditions of satisfaction. Indeed, I will argue later on that the key to understanding intentionality is conditions of satisfaction, but in order to say that, we need a few more items in our apparatus.

4. Causal Self-Referentiality

The most biologically basic intentional phenomena, including perceptual experiences, intentions to do something, and memories, have a peculiar logical feature in their conditions of satisfaction. It is part of the conditions of satisfaction of, for example, my memory that I went on a picnic yesterday, that if I really remember the event, then the event itself must cause my memory of it. If we spell out the conditions of satisfaction of the memory, they are not just that the event occurred, but also that its occurrence caused the very memory that has the occurrence of the event as the rest of its conditions of satisfaction. We can describe this by saying that memories, intentions, and perceptual experiences are all causally self-referential. What this means is that the content of the state itself refers to the state in making a causal requirement. The conditions of satisfaction of the memory itself require that the memory be caused by the event remembered. The conditions of satisfaction of the intention require that the performance of the action represented in the content of the intention requires that that very intention should cause that performance. And so on through other cases.

In this respect, intentions, memories, and perceptual experiences are different from beliefs and desires. We can spell out the difference as follows. If I believe that I went on a picnic yesterday, then the formal structure of my intentional state looks like this:

Believe (I went on a picnic yesterday).

But if I remember that I went on a picnic yesterday then the formal structure of my intentional state looks like this:

> Remember (I went on a picnic yesterday, and my going on a picnic caused this memory).

For states that have the mind-to-world direction of fit we need to distinguish those that are causally self-referential, such as perceptions and memories, from those that are not, such as beliefs. Exactly parallel to this, for states that have the world-to-mind direction of fit we need to distinguish those that are causally self-referential, such as the intention that I have prior to doing something (what I call the "prior intention") and the intention I have while I am actually doing it (what I call the "intention-in-action") from those that are not causally self-referential, such as desires. Also every causally self-referential state with a direction of fit also has a direction of causation. In visual perception, for example, if I see that the cat is on the mat, I see how things really are (and thus achieve mind-to-world direction of fit) only if the cat's being on the mat causes me to see the situation that way (world-to-mind direction of causation). In intentional action, the arrows run the other way. I succeed in intentionally reaching the book on the top shelf (and thus achieve world-to-mind direction of fit) only if my trying, my intention-in-action, causes my success (mind-to-world direction of causation).

The resulting formal relations are so beautiful that I cannot resist setting them out in a chart, where I use the old-fashioned terminology of cognition and volition to name the two families:

	COGNITION			VOLITION		
	Perception	Memory	Belief	Intention in Action	Prior Intention	Desire
Causal Self-Reference	YES	YES	NO	YES	YES	NO
Direction of Fit	↓	↓	↓	↑	↑	↑
Direction of Causation	↑	↑	None	↓	↓	None

5. The Network of Intentionality and the Background of Preintentional Capacities

Intentional states do not in general come in isolated units. If I believe, for example, that it is raining, I cannot just have that belief in isolation. I must believe, for example, that rain consists of drops of water, that these fall out of the sky, that they generally go down and not up, that they make the ground wet, that they come out of clouds in the sky, and so on more or less indefinitely. Of course, someone might have the belief that it is raining and lack some of these other beliefs, but in general it seems that the belief that it is raining is only the belief that it is because of its position in a "*network*" of beliefs and other intentional states. And we can think of the totality of one's intentional states as forming an elaborate interacting network. We can even say that any intentional state only functions, that is it only determines its conditions of satisfaction, relative to the network of which it is a part. If I believe I own a car, I must also believe that cars are modes of transportation, that they are used on streets and highways, that they move about,

that people can get in and out of cars, that cars are a kind of property that can be bought and sold, and so on.

If you follow out the threads in the network, you eventually reach a set of abilities, ways of coping with the world, dispositions, and capacities generally that I collectively call the "Background." For example, if I form the intention to go skiing I can do so only if I take for granted that I have the ability to ski, but the ability to ski is not itself an additional intention, belief, or desire. I hold the controversial thesis that intentional states in general require a background of nonintentional capacities in order to function all.

I have given a very brief sketch of the formal structure of intentionality. We can summarize it as follows. For any intentional state, there is a distinction between the type of state it is, and its content. Where the content is a whole proposition, it will represent states of affairs in the world and it will do this with one of the three directions of fit, mind-to-world, world-to-mind, or null. Intentional states that have a non-null direction of fit are thus representations of their conditions of satisfaction. And given the network of intentionality, even those states that have the null direction of fit, and even those that do not have a whole propositional content, are still largely constituted by states that do have a non-null direction of fit. Thus if I am sorry that I stepped on your foot, I must believe that I did so and wish I had not done so. And if I admire Jimmy Carter I must have a set of beliefs and desires about Jimmy Carter. In general, *intentionality is representation of conditions of satisfaction*. The most biologically basic intentional states, those that relate animals directly to the environment, have a causally self-referential component in their conditions of

satisfaction. Any intentional state can function, that is, it can determine conditions of satisfaction, only because of its position in a network of intentional states and against the background of pre-intentional capacities.

Later on, when I talk about the unconscious in chapter 9, we will see that the network of intentionality, when unconscious, is really a special case of background abilities, the ability to produce conscious intentional phenomena.

The formal structure of the intentionality that I have described is no trivial matter. This is in fact the structure of our conscious life. Indeed, it is the structure of our mental life, both conscious and unconscious. When we come to understand a social situation we are in, when we make up our minds to engage in some course of action, when we perceive the heavens on a starry night, when we suddenly have recollections of our childhood while eating a madeleine—all of these are manifestations of the formal structure that I have been describing. In order to understand our lives, we have to understand the structure of intentionality.

It is important to emphasize that none of this discussion is intended to be phenomenological. We are talking about the logical structure of intentionality. Phenomenology, for the most part, is unable to access this structure.

III. INTENTIONALITY-WITH-A-T AND INTENSIONALITY-WITH-AN-S

You will not understand the current philosophical literature on intentionality unless you see the difference between intentionality-with-a-t and intensionality-with-an-s.

These are often confused, even by professional philosophers. Intentionality-with-a-t, as we have seen, is that

property of the mind by which it is directed at or about or of objects and states of affairs in the world independent of itself. Intensionality-with-an-s is opposed to *extensionality*. It is a property of certain sentences, statements, and other linguistic entities by which they fail to meet certain tests for extensionality. The connection between the two is that many sentences about intentional-with-a-t states are intensional-with-an-s sentences. There are several such tests for extensionality, but the two most famous are the substitution test (sometimes called Leibnitz's Law) and the test of existential inference. Let us consider each of these in order. The substitution test says that whenever two expressions refer to the same thing, you can substitute one for the other without changing the truth value of the statement in which you are making a substitution. Formally we can put this as follows:

1. $[(a=b) \,\&\, Fa] \rightarrow Fb$.
   If a is identical with b and a has property F, then b has property F.
   Thus from
2. Caesar crossed the Rubicon.
   and
3. Caesar is identical with Mark Anthony's best friend.
   we can infer
4. Mark Anthony's best friend crossed the Rubicon.
   For this reason, the occurrence of "Caesar" in 2 is said to be *extensional* with respect to substitutability. But there are sentences in which you cannot make the substitution. Thus from
5. Brutus believes that Caesar crossed the Rubicon.
   and the identity statement 3, we cannot validly infer

6. Brutus believes that Mark Anthony's best friend crossed the Rubicon,
   because Brutus might not believe that Caesar is Mark Anthony's best friend. Such a sentence is said to be *intensional* with respect to the occurrence of Caesar. It fails the test of substitutability.
   The principle of existential inference says that whenever a has the property F, you can validly infer that there exists an object that has the property F.
7. $Fa \rightarrow (\exists x)(Fx)$
   Thus from
8. John lives in Kansas City.
   we can validly infer
9. There is some x such that John lives in x.
   But there are sentences of this form where we cannot validly make the inference. Thus from
10. John is looking for the lost city of Atlantis
    It does not follow that
11. There is some x such that John is looking for x.
    Because the city he is looking for might not even exist.

Sentences such as 10 are said to be intensional, because they fail the test of existential inference.

Notice that both of these intensional sentences are about states that are intentional-with-a-t. This has led some philosophers to mistakenly suppose that there is something essentially intensional about intentionality. But that is a mistake. The reason that sentences about intentional-with-a-t states are often intensional-with-an-s is as follows: the states themselves are representations of their conditions of satisfaction. But sentences about those states are not representations of those conditions of satisfaction, rather they

are representations of their representations. Hence the truth or falsity of such sentences does not depend on how things are in the real world as represented by the original intentional states, but how things are in the world of representations as it exists in the minds of the agents whose intentional states are being represented. Thus when I say Caesar crossed the Rubicon I am talking straight out about Caesar and the Rubicon. But when I say Brutus believed that Caesar crossed the Rubicon, I am talking about Brutus and what is going on inside his head. The truth of what I say depends not on the real world of Caesar and the Rubicon but on what is in Brutus' head that represents Caesar and the Rubicon. Thus I cannot make the substitution unless I have an extra premise to the effect that Brutus would accept it. Analogous remarks apply to the test of existential inference. If I talk about where John actually lives, then I am talking about an actual person and an actual place, but if I talk about what John is looking for, I am talking about an intentional state, trying to find something, whose conditions of satisfaction he is attempting to realize. But he might have that intentional state, he might be looking for something, even if the something he is looking for does not exist. Once again, the fact that the intensional-with-an-s sentence is a representation of a representation explains its intensionality.

The important thing to remember about the distinction between intentionality-with-a-t and intensionality-with-an-s is that there is nothing inherently intensional about intentionality. A statement to the effect that Brutus believes that Caesar crossed the Rubicon is indeed an intensional-with-an-s statement. But the belief itself, Brutus's actual belief, does not thereby become intensional-with-an-s. The

belief itself is as extensional as it can get. It will be true only if both Caesar and the Rubicon exist, (existential inference) and anything identical with Caesar crossed anything identical with the Rubicon (substitutability).

I do not want to give the impression that you understand all there is to understand about intensionality-with-an-s on the basis of the preceding paragraphs. There is much more be said. For more details see my book *Intentionality: An Essay in the Philosophy of Mind.*[2] All I want to do right now is give you enough tools so that you can follow arguments about intensionality-with-an-s and intentionality-with-a-t without making the mistakes that are common in contemporary philosophy.

IV. THE DETERMINATION OF INTENTIONAL CONTENT: TWO ARGUMENTS FOR EXTERNALISM

Most philosophers who write about these issues seem to think that there is a very general question, with an equally general answer, of the form, How is the content of our intentional states determined? The question is supposed to be interpreted as asking not, What is the account of how we came to have these intentional contents and not others? but rather, How are the intentional contents *constituted*? What fact about the intentional state as it exists here and now makes it a desire for water and not a desire for something else? Oddly enough, though these are quite distinct questions, the currently most influential view treats an answer to the first, What is the causal account for our having these intentional states? as providing the answer to the second, What is it about these intentional states that constitutes their having the content they do?

This view, called "externalism," says that intentional content is in large part constituted by the (external) causal relations that the agent has to the external world and not by the (internal) features of the mind / brain.

The view that I have been tacitly assuming throughout this book is a form of internalism. According to internalism, so construed, our intentional contents are entirely a matter of what is inside our heads. Of course they refer to objects and states of affairs in the world. That is what intentionality is for—to relate us to the world by representing its various features. The content that enables an intentional state to refer to one object rather than another is entirely between the ears of the referring subject. Internalism, so construed, has in recent decades been challenged by a series of arguments for the view that mental contents themselves are not in the head, or at least not entirely in the head, but in large part reside in relations between what is going on in the head and the rest of the world. It is important to see that this externalist theory is not merely claiming that our inner mental contents are often caused by external events (both sides agree on that) but rather that the contents themselves are not truly inner but are, at best, a mixture of the inner and outer. If that sounds vague, I am afraid it is, because externalism is a rather vaguely stated thesis. I will now sketch the two best-known arguments for externalism, and this will help to make the doctrine seem less obscure. In order to explain these arguments I need to introduce the notion of indexicality. An indexical sentence or expression refers to some object by indicating the relations in which the object stands to the utterance of the expression itself. So if I say, "I am hungry" and you say "I am hungry" we utter the same sentence with the same

meaning but the utterances have different conditions of satisfaction because of the occurrence of the indexical "I." "I" uttered by me refers to me. "I" uttered by you refers to you. There are lots of forms of indexicality in language: "I," "you," "here," "now," "this," "that," "yesterday," "tomorrow," and "over there," as well as tenses of verbs, are all examples of indexicals.

The First Argument for Externalism: Hilary Putnam and Twin Earth.[3]

You might think that "water" could be defined as a clear, colorless, tasteless liquid found in lakes and streams and coming out of the sky in the form of rain. But, says Hilary Putnam, that does not give the meaning of "water." To see this, imagine a galaxy just like ours, with a planet in it just like our planet, that we will call Twin Earth. On Twin Earth everything is exactly the same as it is on Earth, molecule for molecule, with one exception. What we on Earth call "water" is made of H_2O; what people on Twin Earth call "water" is not H_2O but has a very long chemical formula that we can abbreviate as "XYZ." Now, in 1750, before anybody knew anything about chemical composition, what was in the heads of the Twin Earth people when they used the word "water" was exactly the same as what was in the heads of the Earth people when they use the same word. But all the same, though the contents of the heads were the same, the meanings were different. Meanings cannot be in the head, because the same things are in their heads as are in our heads, but the meanings are different. "Water" on Earth refers to one kind of stuff; "water" on Twin Earth refers to another kind of stuff. The meaning on both Earth

and Twin Earth, says Putnam, is determined by causal relations in which speakers stand to indexically presented substances. "Water" on Earth means whatever has the same structure as this indexically presented stuff. Ditto for Twin Earth. But since the stuffs are different, H_2O in one case, XYZ in the other, the meanings are different. Meanings, concludes Putnam, "just ain't in the head."[4]

What goes for meaning goes for mental content generally. Beliefs employing the expression "water" are different for the people on Twin Earth than for the people on Earth. But if so, it turns out that beliefs cannot be entirely in the head. What is in the head is exactly the same in the two cases, though the beliefs are different.

The Second Argument for Externalism: Tyler Burge and Arthritis[5]

Tyler Burge has a related argument to show that the contents of the mind are at least in part social. Here is how the argument goes. Imagine that Joe goes to see his doctor in Santa Monica. He says "Doctor, I have a pain in my thigh. I believe it is arthritis." We may suppose his doctor answers, "If it is a pain in your thigh, it can't be arthritis. Arthritis is an inflammation of the joints." Now let us keep the condition of Joe exactly the same but imagine that the community is different. Imagine that what is in Joe's head is exactly the same because he is the same person at the same time. But let us imagine that he is not in Santa Monica but in Twin Santa Monica. And imagine that in this community the word "arthritis" is used differently. It is used to name both muscle pains and joint inflammations. Now, in the second case, what is in Joe's brain is exactly the same as the first case, but

it seems that his belief is different. In Santa Monica he holds a false belief that he has arthritis. In Twin Santa Monica he holds a true belief. We cannot report this belief by saying that he believes he has arthritis, because "arthritis" is a word of standard English. In Twin Santa Monica, they do not speak standard English, at least as far as this word is concerned. So we have to invent a word. We can say that in Twin Santa Monica he holds a true belief, the belief that he has tharthritis. Now, and this is the point of the thought experiment, though what is in his head in the two cases is exactly the same (it has to be the same because he is exactly the same person at the same time), all the same there are two different beliefs. There must be two different beliefs because one is true and the other is false, and the same belief cannot be both true and false.

The conclusion is like Putnam's. Just as Putnam showed that meanings are partly constituted by causal relations to the world, so Burge's argument shows that mental contents are partly constituted by social relations with one's community. In both cases we seem to have demonstrated that intentional contents are not internal to the head.

What are we to make of these arguments? I admire the philosophical acumen of their authors, but I think both arguments are fallacious. The basic idea of internalism is that the mind—and by "mind" here we mean what is inside the head—sets conditions that an object must meet in order to be referred to by an expression or other form of thought content. In a classic example, the expression "the Morning Star" sets a condition such that if an object satisfies that condition, the expression can be used literally to refer to the object. Nothing in Putnam's account challenges this

conception. For the traditional idea that a checklist of features is associated with each word—for example, with the word "water" are associated such features as clear, colorless liquid, etc.—Putnam substitutes an indexical definition: "Water is anything identical in structure with what we are now seeing." On our account of the causal self-referentiality of perceptual intentionality, that amounts to saying that water is whatever is identical in structure with the substance causing this very visual experience. But that definition sets a condition that is entirely represented in the contents of the mind. People on Earth are seeing a substance they call "water," and they set a condition that will be satisfied by anything else that is relevantly similar to the stuff they have baptized as "water." For people on Twin Earth we tell exactly the same story. They are seeing a substance they call "water," and they set a condition that will be satisfied by anything else that is relevantly similar. The condition is entirely internal to the contents of the mind. Whether or not a substance satisfies that condition is up to the world and not up to the mind, in exactly the same way that for any other internally set condition, such as being the Morning Star, whether or not an object satisfies that condition is up to the world and not up to the mind. Internalism is a theory about how the mind sets conditions. Objects are referred to if they satisfy those conditions. What conditions are set is up to the mind; whether an object satisfies those conditions is up to the world. I have seen nothing in the externalist criticisms that challenges this basic insight.

In the case of Burge's example, the only difference in Joe's mental states in the two cases is an indexical difference. In both communities he believes:

1. I am having this very pain in my thigh. I believe it is arthritis.
   But he also has a background presupposition that we can express as:
2. I take it for granted that my use of words matches that of my community and where there is a difference I will alter my usage to match the community.
   But an application of 2 to the present case yields:
3. I take it for granted that in my community "arthritis" refers to pains like this and if not I will alter my usage to conform to the community.

There is thus an indexical component involved in any use of a public language. The difference between Joe in the first case and Joe in the second case is that the community is different. In the first case Joe is wrong about 3. Pains like that are not called "arthritis." In the second community he is right. Pains like that are called "arthritis." I cannot see that this example poses any problem whatever for even the most naïve versions of internalism. In response to this objection, Burge has told me (in conversation) that he simply wants to stipulate that Joe has no metalinguistic beliefs about how words are used. Quite so. We need not suppose he has thought about the matter at all. But it is a background assumption behind our social use of words that we share common meanings with other people in our community. When Joe finds that this background assumption is mistaken he does not alter in any way his conception of the nonlinguistic facts—he still has the same pain in the same place—but he alters his linguistic usage. I think Burge is right that we can reasonably suppose that Joe never had any explicit thoughts to the effect that his usage conforms

to the community. But the presupposition of commonality of linguistic usage is a general background assumption, something that is prior to explicit beliefs and thoughts. Our use of language is presumed to conform to the other members of our community, otherwise we could not intend to communicate with them by using a common language.

V. HOW INTERNAL MENTAL CONTENT RELATES AGENTS TO THE WORLD

In order to explain in more depth what is wrong with these objections to internalism, I have to say a little bit about the nature of mental content and how it relates agents to the world. We have already seen that an intentional state sets conditions of satisfaction. So for example, if I have the belief that Socrates drinks water then my belief will be true, and hence satisfied, if and only if Socrates drinks water. The questions we are asking now are, What features constitute the components of the thought that Socrates drinks water, and how do those component elements relate the agent to the total thought and to the external world? In this case let us concentrate our attention on "Socrates" and "water." (I will leave out a discussion of "drinks" because predication raises special problems that go beyond the issues of externalism and internalism.) Everybody agrees that each component, "Socrates" and "water," makes a contribution to the total truth condition of the thought. "Socrates" picks out Socrates and "water" refers to water. Just as associated with the whole sentence is the truth condition that Socrates drinks water, so associated with each of these two components is a condition, a condition that it contributes to the truth condition of the entire

sentence. There are then two sets of questions about the components of the thought. First, how does each element relate to the condition that it determines and second, how does the agent relate to the determination of those conditions? Granted that "Socrates" refers to Socrates and "water" refers to water, how does the agent have to relate to these words in order that he can use them to determine the conditions of satisfaction of the whole thought? The traditional answer, and the answer given by common sense, is that each word sets the condition it does because of its *meaning* and the agent is able to use the words the way he does because he *knows* the meaning of each of the words. And knowing the meaning enables him to use the word in such a way as to introduce the corresponding condition into the truth conditions of the entire sentence.

We can now state the dispute between the internalists and the externalists with a little more precision: both sides agree that words make a contribution to the truth conditions of the entire sentence and both sides agree that there is some condition that the speaker himself must satisfy in order that he can use these words to set the truth conditions in question. The dispute is entirely about the nature of the condition satisfied by the speaker. The question is, Is the condition associated with the word something that is represented in the speaker's mind / brain, or is it something that is in part independent of the speaker's mind / brain? According to the internalist, the condition must be represented in the speaker's head. According to the externalist, the contents of the head are insufficient for successful reference. That is what Putnam meant when he said "Meanings just ain't in the head." The argument given by the externalists is in every case the same: two speakers

could have type-identical contents in the heads and yet mean something different. But the answer given to this claim by the internalists is that in all cases where that is so, it is because there is some indexical component in the head that sets a different condition of satisfaction in the two cases, because it sets the condition relative to the head of the speaker in question. If we suppose, for example, that two identical twins who happen to be identical, as they say, "molecule for molecule," both think the thought "I am hungry" we may suppose that what it is in their heads is type-identical, but all the same they mean something different because twin A is referring to himself and twin B is referring to himself. Indexicality will enable type-identical thoughts in the head to determine different conditions of satisfaction because the conditions of satisfaction, being indexically determined, are fixed relative to the head in question. Thus in the Twin Earth case the people on both Earth and Twin Earth set conditions of satisfaction relative to themselves: What we call "water" is anything type-identical in structure with the stuff that *we* are seeing. But since the "we" in the two cases is different and since the people on Twin Earth are seeing something different from the people on Earth they will have different conditions of satisfaction even though the contents of the head are type-identical. There is nothing in this example to show that meanings are not in the head.

Analogous remarks can be made about Burge's example. Joe has exactly the same thought in the two communities. The thought is "I am having this very pain. I believe it is arthritis." And the Background presupposition is that pains like this are called "arthritis" in my community. But since the community is different in the

two cases, the very same thought will determine different conditions of satisfaction relative to the two communities. In one case Joe has a true belief; in the other case he has a false belief.

Let us return to our original question. If we reject the externalist's claim that intentional content is determined by external causal chains, what then does determine intentional content? Causally speaking, I do not think there is any general answer to this question except to say that our intentional contents are determined by a combination of our life experiences and our innate biological capacities. I have already given a sketch of how an animal's feeling of thirst might be determined by neurobiological processes. If one were to change the example slightly so that I was not just thirsty in general but thirsty for a glass of draught Irish stout, or a 1953 Chateau Lafitte, then the story would become much more complicated. I would have to give an account of how my life experiences have led me to have certain sorts of taste experiences, that I was capable of recalling these in memory and capable of forming desires to repeat these experiences in the future. But if the story has to be more complicated to account for a specific desire, then it would become incredibly complicated if I tried to give an account of how one might have formed an intention with the content that I write the great American novel, marry a Republican, or explain intentionality in a single chapter.

But if we are talking not about the history of our intentional states, but about their *constitution,* for example, what fact about me makes it the case that I have the belief that Caesar crossed the Rubicon, then we have to appeal to the notion of conditions of satisfaction.

Before addressing that question directly, let us take stock of where we are. We began this chapter with three questions:

1. How is intentionality possible at all?
2. How are intentional contents determined?
3. How do intentional states work in detail?

We did not so much answer the first question as remove the need to ask it in that special philosophical tone of voice that makes any answer impossible. We brought it down to earth by transforming it into such questions as, How is it possible for an animal to be thirsty, or hungry, or frightened? Once those questions are answered, the first question is already answered insofar as it is a meaningful question. We postponed the second question until we had answered the third. In passing, I rejected the externalist answer to the second question. I now want to use our results in answering the third question to perform the same sort of maneuver on the second that we did on the first. The question, How is it possible for me to have a belief whose content is that Caesar crossed the Rubicon? is in principle no more difficult to answer than it is to answer the question, How was it possible for me to be thirsty for water? i.e., to have a desire whose content is that I drink water. In both cases the answer is provided by seeing the essential connection between intentionality and conditions of satisfaction. What makes my desire a desire to drink water is that it will be satisfied if and only if I drink water. That is not a psychological remark predicting what will make me feel good, but rather it is the definition of the relevant intentional content. In exactly the same way, what

makes my belief have the content that Caesar crossed the Rubicon is the fact that it will be satisfied if and only if Caesar crossed the Rubicon. The content of the intentional state is exactly that which makes it have the conditions of satisfaction that it does. Those conditions of satisfaction are always represented under aspects. I represent a certain man as Caesar, for example, and not as Anthony's best friend, even though Caesar is identical with Antony's best friend.

But is not this answer to the second question circular? What makes an intentional state have the content it does? Answer: it has the conditions of satisfaction that it does. And what are those conditions of satisfaction? Those determined by the content of the intentional state. And that certainly looks circular. But that is precisely the sort of circularity I am seeking. We do not accept the question on its own terms, but rather reject it and substitute for it an account of how intentionality actually functions. It functions because of a set of very tight connections between intentional content, aspectual shape, and conditions of satisfaction. The next step in nailing this whole account down to the real world is to point to the central role of consciousness. To have an intentional state consciously, for example to think consciously that Caesar crossed the Rubicon, is to be consciously aware of the conditions of satisfaction. To have the same intentional state unconsciously is to have something that is in principle is at least capable of becoming conscious. I will discuss the relation of the conscious and unconscious in detail in chapter 9. For present purposes I want to say only the following. We reject the sense of the third question in which it does not admit of any answer and we substitute for that question an account of how intentional content actually functions. It

actually functions because intentional agents have conscious thoughts where the very identity of the conscious thought is such as to determine that it has certain conditions of satisfaction and not others. Those conditions of satisfaction are represented under some aspects and not others. If you ask, How can a state of my brain have the content that Caesar crossed the Rubicon? it seems an impossibly difficult question. But if you ask, How can my conscious thought "Caesar crossed the Rubicon" have the content that Caesar crossed the Rubicon? Then it is no longer impossible to answer. I know the meanings of the words, I know how they relate to objects and states of affairs in the world and in thinking the whole thought I am aware that it has precisely this condition of satisfaction: Caesar crossed the Rubicon. Once we reject the metaphysical sense of the third question we demystify it by assimilating it to a general account of how intentionality actually functions. And that is all that needs to be said about the constitution of intentional content in general. Beyond that, of course, we need to say a great deal, much of which I have already said, about the network and the background, about the direction of fit and causal self-referentiality, psychological mode, and all the rest of it.

I will spell out the relations between consciousness and intentionality in chapter 9. For the moment, just this: one huge evolutionary advantage of human consciousness is that we can coordinate a large amount of intentionality ("information") simultaneously in a single unified conscious field. Think of the amount of coordinated intentionality ("information processing") when, for example, you drive to work in the morning. Don't just think of the coordination of perception and action. (For example, I am

passing the car on my right. There is a red light ahead.) Think also of the constant accessing of unconscious intentionality. (For example, I will be late for my 9:00 a.m. appointment. Where shall I have lunch? I wonder how the meetings will go.) All of these are intentionalistic representations of the world, and we cope with the world by way of these representations.

VI. CONCLUSION

I said at the beginning of this book that the worst thing we can do is give the reader the impression that she understands something she does not really understand. I do not wish you to get the impression from reading this chapter that now you understand intentionality. I have only scratched the surface of a very large subject. But I do want you to have a certain overall conception of intentionality as representation and I do want you to be able to avoid mistakes that are common in contemporary philosophy. Specifically, you should see the distinction between intentionality-with-a-t and intensionality-with-an-s. You should see the difficulties in the currently orthodox externalist accounts of intentional content, and you should begin to see the connection between intentionality and consciousness, a connection I will explain in detail in chapter 9. Most of all, you should begin to get an idea of how intentionality works as a real feature of the real world, and this understanding will, I hope, enable you to avoid being intimidated into thinking there is some deep mystery about intrinsic or original intentionality that defies any natural explanation.