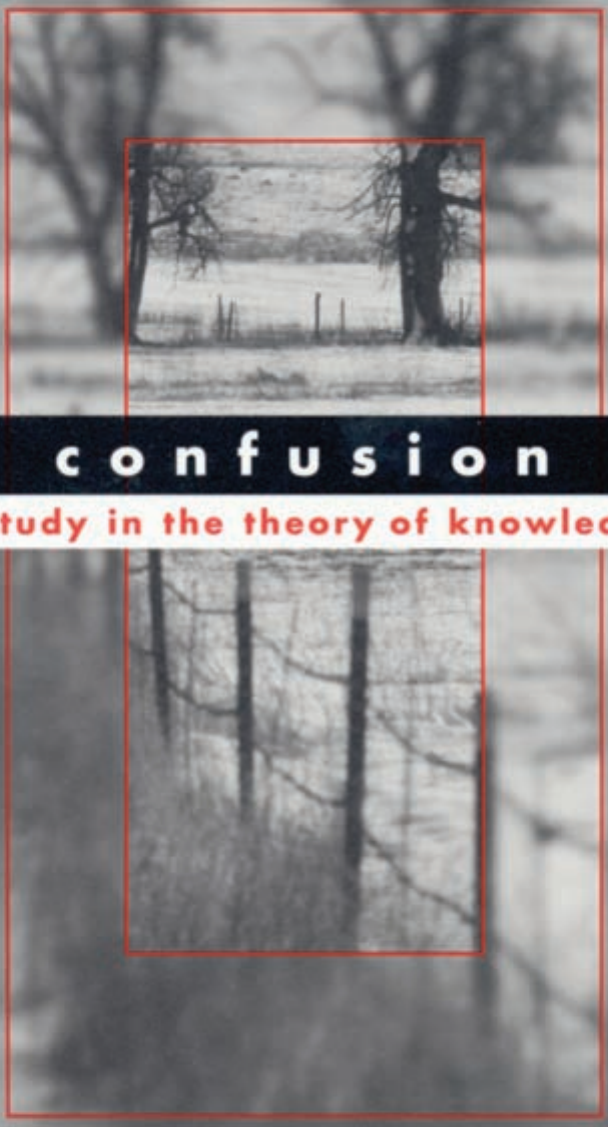
JOSEPH L. CAMP, JR.



confusion

a study in the theory of knowledge

Confusion

Joseph L. Camp, Jr.

Confusion

A Study in the Theory of Knowledge

Harvard University Press
Cambridge, Massachusetts, and London, England | 2002

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Camp, Joseph L.
Confusion : a study in the theory of knowledge / Joseph L. Camp, Jr.
p. cm.
Includes bibliographical references (p.) and index.
ISBN 0-674-00620-8 (alk. paper)
1. Errors. 2. Knowledge, Theory of. I Title.

BD171 .C35 2001
165—dc21 2001239773

I have worked for about twenty years on the problem of this book, and on the larger problem—the relation between truth and reason—of which it is a part. Half a dozen times I would have given the thing up but for Tamara, who was the wind beneath my wings.

Preface

Before she fell ill with the brain cancer that ended her life, Tamara Horowitz read every draft of every paragraph that eventually entered this book, as well as twice that number of paragraphs no longer in existence. She squeezed out a lot of the hot air, and is not to blame for such hot air as remains. Tamara also thought of several arguments I needed but could not supply; I have incorporated versions of all of them, especially in chapters 2 and 9. Several friends read carefully through a penultimate draft, and gave me valuable criticisms and suggested revisions—and some more arguments I needed but could not supply. The list: Anil Gupta, John MacFarlane, Walter Edelberg, Richard Gale, Allen Hazen, Mark Wilson. I hope I'm not forgetting anybody. I presented portions of the material in this book, and related material, in graduate courses at the University of Pittsburgh over several years. Many students in those classes gave me criticisms and suggested new points of view; I won't try to list everybody. As Ron Haver once remarked in a similar context, they know who they are.

Anonymous referees for the Harvard University Press gave me criticisms of the penultimate draft and suggestions for revising it, and John MacFarlane's Fall 2000 graduate class at the University of California, Berkeley, transmitted various questions and criticisms to me via John, all of which were clarifying. It was especially useful to have thoughtful reactions from readers who, unlike my friends and students, began with no idea where I was coming from or where I was trying to go. (Some of them may feel they ended, as well as began, with no idea where I was coming from, though I hope not.)

One needs a benchmark. One needs to know when a consideration lacks subtlety enough to be worth raising; one needs to know where the minimum acceptable level of clarity lies, below which one may not drift, and

one needs to know where the optimum level lies; one needs to know when rigor has been sacrificed unacceptably to chattiness, or chattiness sacrificed unacceptably to rigor. These are the tools of the art, and one needs to be reminded, by example, of their right use. Nuel Belnap has been my benchmark.

I would not have been a philosopher, or else would not have remained one, were it not for the intervention of two people, friends and teachers both, at some time in the past. When I thought philosophy was interesting, Panayot Butchvarov showed me it was important; and when I thought philosophy was easy, Wilfrid Sellars showed me it was hard.

Contents

I

MATERIAL FALSITY

1

Thinking One Thing Is Another

After working late one night you walk to your car to drive home. Your car is parked in a dimly lit parking lot. There are several other cars in the lot. You are tired, and your mind is not really on what you are doing. You are on automatic pilot, thinking about what you will do when you get home at long last. You go up to what you think is your car, and start to insert the key into the door lock. It won't go in. This annoys you, and for a moment you keep wiggling the key in the lock, wondering why it is so balky. Then you notice the stuffed panda in the back seat. You do not carry a stuffed panda in your back seat. It isn't your car. You look around and quickly spot your car across the lot, where you now suddenly recall leaving it. The car you have been trying to unlock is of the same make and model, and near enough in color to be almost indistinguishable in poor light.

For a few seconds you thought this car was your car. You confused the two. You took one for the other. Not all confusion is of this kind—taking one thing to be another. Sometimes a person is temporarily, or even permanently, the victim of muddled thought on a much larger scale. Think about what it is like to be stunned by a blow to the head, or to be very drunk, or to be drugged, or to be wakened suddenly from a deep sleep—if you have ever had any of those experiences. One speaks of "confusion" in these cases too, although the cognitive abnormality is more diffuse, and often more debilitating, than it is when one merely mistakes one thing for another. In order to narrow the application of the term "confusion" to the taking-one-thing-for-another variety, I sometimes use the term of art "ontological confusion," meaning confusing objects as opposed to being generally foggy and muddle-headed. In this essay I will use "confusion" always to mean "ontological confusion," just to save ink.

Confusions sometimes are cleared up quickly, as with the car in the parking lot. But they can be more persistent. I once met a man, spent a bit

of time with him, and some years later met, in similar circumstances, a different person with a similar name and a similar appearance. I confused the two and did not manage to correct the confusion for several more years. Most people I have asked say they have had such experiences. Some people have had many. It seems likely that some, perhaps most, confusions of this kind are never corrected, since they involve strangers, or passing cars, or city buses, or squirrels, or roadside diners in northern Georgia.

It is perfectly obvious that when you confuse the car in front of you with your own car, you have made a mistake. Indeed, you have made the paradigmatic "mis-take." It is perfectly obvious that you merit intellectual criticism—some kind of intellectual criticism. It is perfectly obvious that you would be better off not being confused. And it is perfectly obvious that seeing the stuffed panda in the back seat results in your confusion abating. But it is not perfectly obvious what kind of intellectual criticism you deserve for making your mistake, or why your mistake dissolves in the face of that panda.

The discipline of epistemology, the "theory of knowledge," is devoted to making philosophical sense of the various mistakes people can make. Epistemologists try to describe these mistakes in such a way that one can say very precisely why the mistakes are defects, why it is a bad thing to suffer one of them, and why the sufferer deserves intellectual criticism. The ultimate aim is to explain what it takes to correct the mistake, and what it takes to avoid making it. But nobody in the world knows how to do any of these things for the type of mistake you made in the parking lot. Nobody, including me. I have come to have some opinions about how the epistemology of confusion should be approached, and some partial arguments for those opinions. This book is a summary of both the opinions and the partial arguments.[1]

Anyone who has never thought about confusion as a distinct epistemological topic (like me, not so long ago) would do well to work up a sense of bewilderment at what to make of it. The most bewildering aspects of the phenomenon will take us a while to dig out from down inside, but even the surface is weird. Picture yourself just after you first tried to put your key in the door lock of the wrong car. You thought the car before you was your car. Suppose you still owed five hundred dollars on your car. And suppose you were inclined to obsess about the matter. While you were jiggling the key in the door lock, you had the thought, and said aloud: "I owe five hundred dollars on this car." Which car were you thinking and talking about? And were you right or wrong? Here are some options:

Option 1. You are using the expression "this car" to refer to your car; you are not using the expression to refer to the car in front of you. Your statement "I owe five hundred dollars on this car" is true, because you do owe five hundred dollars on your car. You are having, and vocalizing, a thought about your car *and just it* while messing around with a different car. Perceptible features of the car in front of you, taken together with other psychologically relevant factors (fatigue, distractedness, and so on), have caused you to think your car is right there before you. Your use of the expression "this car" is related psychocausally to the car in front of you, but not semantically (except negatively—it is no more "about" or "of" that car than it is about or of the World Trade Center).

Option 2. You are using the expression "this car" to refer to the car before you, and just it. Your statement "I owe five hundred dollars on this car" is false. You believe, of the car before your eyes, that it possesses many properties that it does not possess, but that your car does possess. The visual similarity between the car in front of you and your car, together with other psychologically relevant factors (distractedness, fatigue), has led you to attribute various properties to the wrong subject.

Option 3. You are using the expression "this car" to refer to your car, *and* to refer to the car before you, at the same time. The thought you verbalize, "I owe five hundred dollars on this car," is about two objects at once, not about any single object *and just it*. This can be seen from the fact that the rationale for (1) and the rationale for (2) are both compelling. The way to break the impasse is to suppress the unique-reference clauses in (1) and (2), and then combine them.

To see how to apply the concepts of truth and falsehood in the presence of "multiple reference," play a little game. Imagine that you, now, could have a chat with you, then. Imagine that you, now, ask a variety of questions, such as "Does this car have a windshield," or "Do cars of this make usually get good gas mileage," or "Do you owe any money on this car." You, now, listen to the answers you, then, give to these questions. Sometimes the statement you, then, make is true regardless of how the expression "this car" is understood—whether it is understood as referring to your car (the one you owned then) or to the car you, then, are standing in front of. For instance, the statement "This car has a windshield" is a true statement whether by "this car" you, then, are taken to mean the car in front of you, then, or the car owned by you, then. So there is no harm done if you, now, accept this statement. The statement conveys no misinformation about either car. Similarly, the statement "This car has a rocket

engine" is false regardless of which way the expression "this car" is understood. So there is no harm done if you, now, take that statement to be false if it happens to be uttered by you, then. But the statement "I owe five hundred dollars on this car" is a different matter entirely. If you, now, take it to be true, then you are taking as true a statement that conveys misinformation about one of the cars, whereas if you take it to be false, then you are effectively accepting a statement (the negation of the original) which conveys misinformation about another one of the cars. Either way, there is harm done. So you, now, ought not regard the statement "I owe five hundred dollars on this car" as made by you, then, as a true statement or as a false statement. This "no harm done" policy for deciding whether a statement made by you, then, is true, or false, or neither is a form of "supervaluation."[2]

Back to you, then, in the parking lot. Options (1), (2), and (3) all attempt to answer the questions "What car are you thinking about" and "What car are you talking about" without tinkering with one of the presuppositions such questions would ordinarily have. The presupposition might be called a "realist" presupposition: you use such expressions as "this car" to refer to objects in the real world, and as a result the this-car thoughts you have and the this-car statements you make are about objects in the real world. But not all singular terms refer to objects in the real world, and not all thoughts and statements are about objects in the real world. When a Norseman said (in Norse) "The Great Wolf Fenrir is coughing and sputtering out in the big sea today," he was not talking about anything real, although he was "talking about something"—as opposed to talking about nothing.

The Great Wolf Fenrir is an unreal, mythical object. The Norseman and his kin attributed a range of wind and weather phenomena to the huffing and puffing of Fenrir. These phenomena actually are caused by a succession of complex sea-and-air events modern meteorology has identified more or less accurately. There is no single entity, no real-world Fenrir, out there causing the phenomena the Norseman observed. To have some compact jargon, let's say that the Norseman had Fenrir-thoughts and made Fenrir-statements *due to encounters with* many different unseen oceanic weather systems. "Being due to encounters with" is a causal, not a semantic, relation between a thought and the world. A Norseman's Fenrir-thought was not a thought "about" whichever weather system was beyond the horizon kicking up trouble. It was about Fenrir.

Option 4. The best way to look at the odd cognitive state you are in

while standing there in the parking lot is to assimilate your condition to the condition of the Norseman. You are not using the expression "this car" to refer to your car, nor to refer to the car before you, nor to refer to both cars at once. You are not using it to refer, singly or multiply, to any real object(s). You have come to accept a mythical story, and in that story there is a "your car" character. You hold to this myth only briefly, unlike the Norseman. And very little of your life or your view of the world is affected by your myth, again unlike the Norseman. Before you fell into this myth-ridden state, you were able to have thoughts about a real object which you called your car. Soon you will be able to have such thoughts again. That might make someone believe that for those confused seconds in the parking lot, you must also be able to think about your car, maybe "blended" with something else. But it is more reasonable to regard you as having lapsed into myth without introducing new words for the characters populating that myth. Your this-car-thoughts are due to encounters with your car, and encounters with the car before you. But that is a psychocausal remark about your thoughts, not a semantic remark. From a semantic point of view, you have shifted the reference of certain of your expressions away from the real world to the characters of a myth. In this myth of yours, the your-car character is in front of you, you are trying to open its door, and you owe five hundred dollars on it. When you say "I owe five hundred dollars on this car" you are speaking the language of this myth, using the expression "this car" to refer (singly) to the your-car character in the myth.[3]

The question of the truth or falsehood of your this-car statements has a two-tiered answer, at least at an intuitive, common sense level. We all know how to "enter into" a myth. One would say to the Norseman (in Norse, as usual) "Yes indeed. Old Fenrir is coughing and hacking a powerful lot today. Sure hope he doesn't take one of his sneezing fits." It is fairly easy to master the concepts "true-in-the-myth" and "false-in-the-myth." Those could be our choices for the meaning of "true" and "false." Given those choices, your statement "I owe five hundred dollars on this car" is true. The alternative meaning to give to "true" and "false" is "true of reality" and "false of reality." Some statements of a myth can be assayed as true or false of reality, but only when the statements do not essentially involve terms referring to merely mythic characters. For example, it may be part of the Fenrir myth, as spun by a typical bard, that out in the middle sea there are many icebergs. That part of the myth is "true of reality." But the part about Fenrir standing high above the highest iceberg cannot be

assayed as true or false "of reality." Of course to say we know how to make these sorts of judgments as a matter of common sense is not to say that we have a good philosophical understanding of what we are saying. But it is a start.

Let's add one more option to our list, though not with the implication that the list is complete.

Option 5. The concept of reference is an everyday concept, and we all know how to apply it in run-of-the-mill circumstances. But you are not in run-of-the-mill circumstances in that parking lot. Trying to apply the concept of reference to your expression "this car" is like trying to apply the concept "impeachable offense" to Peter the Great as if he were some high U.S. official. The best one can do is say that if a high federal official in the United States were to perform an act very similar to a given act performed by Peter the Great (such as murdering a bunch of soldiers for siding with his sister), that official would thereby commit an impeachable offense. That might be an instructive analogy for a historian to draw. But the historian would have no right to rely on her analogy as an argument that Peter the Great actually did commit an impeachable offense, as though there had been such a thing in his circumstances—with only a bit of ingenuity being required to figure out which offenses they were.

Similarly, one can analogize between the way you use your words in the parking lot, and other ways you might use your words. In fact there are two instructive analogies. You might be unconfused; you might use the expression "this car" to refer to the car in front of you fully realizing it was not your car. Moreover, you might believe the car in front of you shared many properties with your car it does not in fact share (indeed, for some odd reason, you might believe you owe five hundred dollars on it). On the other hand, you might be unconfused and use the expression "this car" to refer to your own car, even though as you spoke you happened to be standing near the other car, looking at it distractedly. These analogies are interesting because they bring out two different ways your confused remarks are liable to mislead others. But just as the historian has no right to infer from her analogy that Peter the Great committed impeachable offenses, we have no right to infer from the analogies that when you actually use the expression "this car" as you stand in the parking lot, it refers to one or both of the cars. Instead, we should refuse to apply the concept of reference at all.

To suggest that you are referring to something runs counter to a core intuition concerning reference, an intuition captured by Peter Strawson's

form of the theory of descriptions. The intuition is that one refers to objects as part of a larger package. The concept of reference is closely linked to the concept of predication (including relational predication). When a person purports to refer, purports to have a referential thought or to make a referential statement, the person has not really referred unless, as a necessary condition, the person has ascribed a property to the referent or—more generally—placed the referent in a relational nexus with other objects. If the person has not singled an object out, then she cannot have accomplished this whole job. Having more than one candidate for the role of subject-of-predication is just as bad as having none. In the parking lot you fail to single out a unique subject of predication, so you are not referring. Of course you are doing something very similar to things you might do when you are referring, just as Peter the Great did things very similar to things a high official of the United States might do in the course of committing an impeachable offense.

Options (1) through (5) are pairwise inconsistent, and plausible taken singly. I have believed each at one time or another and have been tempted to believe all at once, self-contradiction seeming a small price to pay for just being done with it. The reason it is hard to decide among the five—and one could add other options—is that all of them make semantic claims, but the rationale for each of the five semantic claims is filled in too quickly, too schematically.

When I use the adjective "semantic," as in "semantic claim," and when I use such cognate turns of phrase as "a semantics for such-and-such bit of language," I have in mind just one of the several meanings the word "semantic" has in logic, linguistics, and the philosophy of language. As I use the term, "semantics" concerns itself with explaining what makes a piece of reasoning "logical." The paradigm example of semantics in this sense is an explanation of the semantic property of *validity,* a property of arguments, or *following-from,* a relation the conclusion of an argument can bear to the premises, where the explanation is formulated in terms of properties of the component sentences of the argument—for instance, the properties *truth* and *falsehood*. The explanation might go on to assert that the truth and falsehood of sentences depends upon how certain other properties conspire together—properties of words or phrases. Often, though not always, these properties are language/world relations. Familiar choices are *refers to* and *applies to* (or *is true of*). Logicians sometimes use the terms "semantic" and "semantics" in this sense; theirs is the usage I follow.[4]

When I sketched the five options for understanding what happens in the parking lot, I made free use of such semantic concepts as truth and reference, although in the case of Option (5) the point was negative: one might wish to appeal to a semantics which makes no use of reference. My description of the five options was not entirely semantic; there was some psychological speculation as well, for instance, the remark in connection with Option (2) that "the visual similarity between the car in front of you and your car, together with other psychologically relevant factors (distractedness, fatigue), has led you to attribute various properties to the wrong subject."

In order to decide on a criterion for the validity of arguments, it is necessary to be very clear in one's mind about various wider epistemological concerns. Here is a simple example to illustrate the point: suppose we want a validity criterion for arguments, and the only logical terms we care about are the connectives "not" and "and." That is, we want to decide on a "propositional logic," or a "sentential logic."

Assume that we are wedded to the strategy of explaining validity as follows: an argument is valid precisely when *it preserves truth no matter what, and also preserves nonfalsehood no matter what*. That is, we believe the conclusion of an argument *follows from* the premises exactly when (a) if all the premises are true, so is the conclusion; and (b) if none of the premises is false, neither is the conclusion. Perhaps we are wedded to this strategy for explaining validity because we have a certain conception of the goal of any rational inquirer: an inquirer should seek to believe what is true and disbelieve what is false. So the very best possible inferences never will fritter away the truth of one's premises (if one is fortunate enough to have true premises to work with), and the very best possible inferences never will spit out a false conclusion when there was no falsehood anywhere among one's premises (if one is fortunate enough to have no false premises to work with). Conditions (a) and (b) might collapse into one, but need not.

Our strategy for explaining validity does not compel us to choose any one particular semantics. For instance, we might adopt a "classical," two-valued semantics. When a sentence "A" has the value Truth (somehow or other explained), the negation of the sentence "not A" has the value Falsehood. When a sentence has the value Falsehood, its negation has the value Truth. A conjunction "A and B" has the value Truth if both "A" and "B" do; otherwise it has the value Falsehood. This seems sensible. It is easy to give a precise statement of a concept of validity in this setting. Call sen-

tences containing no logical terms "atomic." Let "A_1, A_2, . . . , A_n; therefore B" be an argument. The argument is a *valid* one precisely when the following occurs: for any distribution of the values Truth and Falsehood to atomic sentences, if each of the premises A_1, A_2, . . . , A_n has the value Truth, then the conclusion B also has the value Truth.

Now here is a worry. There are meaningless sentences which are grammatical, in some intuitive sense. You can tack them together with meaningful sentences in compounds, and the result seems grammatical. For instance, you can conjoin the sentence "Bill Clinton likes jogging" with the sentence "Powers of ten loathe cranberries," obtaining the conjunctive sentence "Bill Clinton likes jogging and powers of ten loathe cranberries," and then ask some interesting questions about what you have done. Does the presence of a meaningless conjunct infect the whole conjunction with meaninglessness? (Quick yet plausible answer: sure does). Does the same thing happen with disjunction? (Same quick yet plausible answer). Should one expect a rigorous thinker to reason validly even when one of these meaningless but grammatical sentences occurs in her thought or speech (or occurs as a clause of another sentence)?

This question has no quick answer. But maybe, upon reflection, we will decide the answer is yes. Here (briefly) are two different reasons we might have:

(1) We use *reductio* arguments to show that the denial of something we want to prove has patently impossible logical consequences. The fact that our premise—the denial of what is to be proved—is itself a necessary falsehood does not prevent us from constructing a valid argument in which it occurs. Perhaps in a similar spirit we should allow ourselves the right to give something like *reductio* proofs where meaningless components are present, as a way to pile up absurdities and thereby persuade some real knucklehead to stop saying such things.

(2) In the course of scientific advance, theses that at one time are meaningless or awfully near to meaningless come to have a meaning at some later time. Is there, sometimes, a transition period when investigators want to reason out whether to accept the hitherto meaningless theses, employing arguments which contain those very theses and related, equally meaningless ones? Imagine someone—not too long ago—asking herself "Could the space of the universe be limited in volume, but without boundaries? That barely makes sense, but let me think how it might happen." Perhaps in our role as logicians we should leave investigators the right to reason in

partly meaningless terms if they choose, tell them the rules for doing so "logically," and let them decide if that is the way to go.

Suppose we are moved by (1) or (2) or some other consideration to admit grammatical-but-meaningless sentences and clauses as proper elements of inference. Our classical, two-valued semantics does not have a value "meaningless" to go alongside "true" and "false." But the so-called Weak Kleene Scheme does, and that might be the semantics we choose. According to this semantics, negations are false when the negated sentence is true, true when it is false, and meaningless when the negated sentence is meaningless. A conjunction is meaningless if it has even one meaningless conjunct, but in the absence of meaningless conjuncts a conjunction is evaluated exactly as in the classical, two-valued semantics. There are some interesting results. For example, conjunction elimination (the form of argument "A and B, therefore A") is invalid. That makes some sense: A might be false and B meaningless. Then the conjunction would be meaningless on the assumption that the meaninglessness of a conjunct infects a conjunction. So the argument "A and B, therefore A" leads from a single meaningless premise to a false conclusion, violating our strategic rule that one cannot validly infer falsehood without a false premise. Inquirers might not think of that without the guidance of a semantics.[5]

Which is the better semantic characterization of "not/and" reasoning, the classical two-valued semantics or the Weak Kleene Scheme? There is no way to answer the question without deciding whether thought involving meaningless components can and should be held to standards of rigor. And there is no way to decide that without taking a stand on broad epistemological positions such as (1) and (2).[6]

My descriptions of Options (1)–(5) were too sketchy. I did not spell out the broader epistemological concerns that arise in connection with attributions of confusion. Just as one could not know whether to prefer a classical two-valued semantics or the Weak Kleene Scheme without reflecting on certain broad epistemological concerns, one cannot make a reasonable choice among Options (1) through (5) without reflecting on what else is at stake, epistemologically, when one attributes confusion. In later chapters I will bring out some of these broader concerns. I will end up plumping for a version of Option (5). A case can be made for a version of Option (4) as well, but that is an exercise for another occasion.

But before we push ahead with our attempt to give a useful analysis of confusion, we should take a look back. The parking lot story is short and simple. In order to get a grip on a philosophical concept, one needs just

such short and simple illustrations. One also needs at least a little understanding of the way a particular concept came to matter in philosophy. It appears that the concept of confusion first came to matter in the course of Hellenistic era debate between Stoics and their critics. Being no classical scholar at all, my Cliff's Notes version goes this way: Stoic epistemology held that all distinct objects are perceptibly distinct. If two things really are two, you can perceive a difference, at least in principle. The obvious sort of reply first would lay out a story like the parking lot story, and then would say "so there." The equally obvious counterreply would be "but in principle, if only you were wise enough, you would know how to tell the cars apart, although in fact you cannot." And so the topic of confusion was on the table.

In the early modern period in philosophy there was a revival of interest in debates from the Hellenistic period. Elements of Stoic and Skeptic epistemology and metaphysics show up in the thought of several early modern philosophers, although the path of inheritance sometimes is hard to trace. Descartes is a case in point. A helpful nutshell description of Descartes the philosopher would be: an intellectualized, Christianized Stoic. It is possible, he believed, for one to have perceptions which exactly discriminate one object from another, whenever they really are distinct objects, with no risk of "parking lot" mixups, but these perceptions are intellectual, not sensory, and in order to suppress the risk of confusion one needs to know there is a Christian deity. Poor Stoics—wedded to the senses and pagan both.

The concept of confusion did not remain at the front of philosophers' minds much longer than Cartesianism remained in vogue. This is not mainly a book in the history of philosophy, but when one's topic came, mattered, and then vanished from the scene, it's worth a chapter.

2

A Little History

The examples of confusion I have mentioned so far involve a person thinking one particular is another particular—where "particulars" are contrasted with "universals." Cars and people and diners in Georgia are particulars. But it is also possible to confuse universals, either properties or relations. Confusions of universals almost never arise from simple misperceptions, or even from transient psychological aberration induced by fatigue or anxiety or absent-mindedness. It is common, though, for someone to confuse two properties or relations because he or she lacks conceptual sophistication. The person does not have enough "theoretical" understanding of the world to tell the two properties or relations apart. Here is an imaginary example, though one that probably closely resembles the condition of many real people.

Marvin knows nothing about physics. He tells which things are cooler than which others by feeling them with his hands or by walking on them with his bare feet. Therefore he cannot discriminate differences in temperature from differences in specific heat. To put it roughly, when heat is being sucked out of the soles of his feet or the palms of his hand at two different rates by two different objects, he cannot tell whether this is happening because the objects differ in temperature, or because the objects differ in heat capacity. And he does not have the scientific sophistication required to realize that there are these two different possible explanations for the fact that one object feels cooler to him than the other. So it never occurs to him that he ought to check in some way to see which mechanism is at work, or if both are. He says: "Stone floors generally are cooler than wood floors"; "Iron ashtrays generally are cooler than books, even when they are in the same room"; and "Boards you leave in the shade generally are cooler than boards you leave in the sun."[1]

Marvin routinely confuses the relation *X having a lower temperature*

than Y with the relation *X having a higher specific heat than Y*. We might describe Marvin's plight by saying, "Marvin thinks (positive) specific heat differences are the same thing as (negative) temperature differences." But when a person confuses objects one ascribes, such as properties or relations, rather than objects one ascribes things to, such as cars or people, the use of the identificatory "is" in the construction "thinks A is B" is not as natural as certain other turns of phrase. More likely, we would say "Marvin does not know the difference between specific heat difference and temperature difference."[2]

One striking fact about Marvin is that his confusion is not something he falls into as his life goes on; it is something he starts with. Curing him of it will take more than hauling him around a parking lot demanding that he pay better attention to pandas and parked cars. It will take a bit of scientific education. It is tempting to wonder how far one could get with a philosophy that took Marvin as the paradigm case of humanity. Maybe we all start with a "common sense" conception of the world so theoretically unsophisticated that, like Marvin, we confuse one property or relation with another property or relation, except that we do this for a very large number of properties and relations. Maybe we do it for all, or nearly all, of them. And maybe as a result we are in error, in a sense, all the time. The only cure for this pervasive error, our philosophy would contend, is an improved theoretical grasp of the reality we interact with. The route to this improvement is scientific education.

René Descartes defended exactly this kind of "pervasive confusion of universals" conception of common sense. Or rather, he defended a special case of it. To simplify formulation, let's use the jargon "due to an encounter with" just as we used it in connection with the "myth" conception of confusion, Option (4) of the previous chapter. Descartes believed two different things about the thoughts comprising prescientific common sense about the surrounding physical world ("common sense," for short).

First, he believed that these common sense thoughts are due to (a) encounters with micromechanical events in the physical world, and (b) encounters with sensory events in one's own mind. Sometimes there is the additional complicating factor of "silly interpretation" imposed on these already doubly-caused thoughts, but we can ignore that. Actually, we need to walk softly here; the "due to encounters with" jargon is causal, in our sense of "causal"—efficient causation. It is not explicit in any text that Descartes believed sensory events are partial *efficient* causes of commonsensical thoughts. I think he did believe it, but this is not the place to ar-

gue the matter.[3] Let's stipulate that he did. If he did not, then the meaning of the language "due to an encounter with" must be expanded to capture his intent.

Second, Descartes believed that common sense thoughts are confused. Any given thought confuses a complicated micromechanical property with a sensory-mental property. This alleged confusion is different from Marvin's in an important respect: Marvin confuses two different physical relations. Descartes thought we all confuse certain physical properties (and relations) with certain "matched" *mental* properties (and relations). Here is an illustrative example.

When you place your hand on snow, you perceive its coldness. There is a property in snow that deserves to be called its "coldness." Snow, like all tangible matter, has a physical microstructure. On a microscopic or submicroscopic scale it is organized like a marvelously complex machine. The real, physical coldness in snow is some aspect of this microstructural organization of the snow. Descartes suspected the coldness in snow was a relatively low overall speed of the microcomponents' motion (low relative to the overall speed of the microcomponents in boiling water, for example), but not having taken the "statistical turn," he was unable to guess at how to build some real science on his hypothesis. When you perceive the coldness of the snow, you are having an experience, a thought in his wide sense of "thought," that is due in part to your encounter with this real, physical aspect of the real, physical microstructure of the snow.

Moreover, your thought is in part *about* this micromechanical property. In general, when you think about the coldness of snow or talk about the coldness of snow, you are thinking or talking (in part) about this real, physical coldness. It simply is not clear whether Descartes meant to identify this "aboutness"—a relation we would think of as semantic—with the causal "due to an encounter with." Since nobody at that time made the distinction, it would be mere speculation either way. One can speculate that if someone traveled back in a time machine and explained the question fully, Descartes would have opted for a "naturalistic" semantics, holding that what it is for a thought to "be about" a micromechanical property is just that the thought "be due to an encounter with" the property. And one can speculate to the contrary. Let's do neither.

The micromechanical property, "physical coldness," is not the only thing your thought is about, just as it is not the only causal source of your thought. Your thought also is partly about a type of sensation, a sensory

property. It is about the very sensory property that characterizes your sense-experience when you encounter the physical coldness of the snow. You encounter this sensory property as well, and your thought "This snow is cold" is due to your encounter with the sensory property (duly instantiated by your mind). The sensory property is a "way things can feel"—the cold way.

Descartes did not talk much about the reference of words. To get the effect of saying that you, the common sense thinker, confuse physical coldness with the cold way of feeling, he spoke in terms of ideas. Where we would say, "Your word 'cold' refers at once to physical coldness and to the cold way of feeling," Descartes would have said, "Your idea of cold is confused. You refer it to both physical coldness and the cold way of feeling." Setting aside differences of philosophical jargon, Descartes was accepting Option (3) of Chapter 1: confusion involves multiple aboutness or multiple reference. He should, therefore, have been much more cautious with his use of the term "error." He should have said that much of common sense is in error in the sense of "untrue," not in the sense of "false." And he should have said the confusion of common sense thoughts is a source of untruth, not a source of falsehood. What he did say was that having an idea like the common sense idea of cold is material for false judgment, so that such common sense ideas may be called materially false.[4]

This confusion charge was a hard sell. Here is Antoine Arnauld, no dummy at all, not getting it—and rather huffily explaining why: "For what is the idea of cold? Cold itself in so far as it exists objectively in the intellect . . . What does the idea of cold, which you say is materially false, exhibit to your mind? A privation? Then it is true. A positive entity? Then it is not the idea of cold."[5]

The trouble was that Descartes and Arnauld had to talk in the idiom of "ideas"—objects in the mind—rather than in the idiom of linguistic items referring to things in the world. A thought got to be "about" some worldly object by involving an idea, an idea that *was* the object but existing intramentally.[6] This was a way of buying aboutness on the cheap: aboutness is identity. But it was an unforgiving doctrine. A linguistic item can refer to just about anything; maybe even several things at once. An intramental entity, it would seem, can be identical with only one thing—itself. Descartes really needed a logic in which identity is permitted to branch, to lack transitivity.[7] But that was no easier to swallow then than now. Arnauld was making the obvious point that if cold-in-your-mind just

"is" cold, and it is the very nature of cold to be an absence of motion, then your idea of cold, cold-in-your-mind, is an absence, or privation. He had Descartes dead to rights.

The remarkable thing is that Descartes, who thought in the same terms, was not thereby prevented from seeing that there was something important one could mean by "my idea of cold is two things at once." But he knew that it would take some fancy footwork to say it without sounding raving mad. It is not clear that he succeeded. With hindsight, from a comparatively semantically sophisticated vantage point, we can tell what he meant. Eventually Arnauld came to believe him, too, though I do not know what changed his mind. Here is Margaret Wilson's translation of Descartes' struggling reply to Arnauld's perfectly sensible complaint: "it often happens in obscure and confused ideas, among which those of heat and cold are included, that they are referred to something other than that of which they are truly the ideas. Thus, if cold is only a privation, the idea of cold is not cold itself, insofar as it is objectively in the understanding, but something else *which I wrongly take for* this privation; that is, a certain sensation which has no being outside the understanding" (emphasis mine).[8]

It was a good plan to start off using the concept of reference, but he simply could not keep it up. What he ends up saying implies that one's idea of cold is "truly" an idea of the privation cold (the physical absence of motion), "referred to" the sensation cold (the cold way of feeling), which therefore is one's idea of cold (as a result of a "wrong taking").

Descartes did not propose theories on the borderline between coherence and incoherence just for fun. In this instance he did it to reconcile two plain facts with each other, and with his mechanism:

(1) It seems that people can fully understand what cold is without reference to the purely spatial and quantitative properties and relations of the micromechanical machine-components posited by mechanism to be the working parts of physical objects.

(2) It seems that coldness is in physical objects, and varies according to the influence of physical objects on one another.

Prima facie, fact (1) is evidence that cold is not a mechanical property. But as mechanism would have it, all real physical properties of real physical objects are mechanical properties.[9] But, again prima facie, fact (2) is evidence that coldness is a real physical property of real physical objects. So taken together, (1) and (2) are prima facie grounds for rejecting mechanism. Various mechanist reactions met with varying success. Locke's much

later suggestion that cold is a mechanically based power to cause sensation had a long run of popularity, with the old idea that (2) reports a universally experienced illusion probably coming in second. Descartes often is portrayed as belonging to the "illusion" camp—cold is to be found only as an inner object of sense, but people foolishly project it onto worldly things. This plainly is not what he was telling Arnauld. He was telling Arnauld that cold is both a physical state, a privation (of motion), and a sensation-type. Force him to take the linguistic turn a few centuries early, and you get the following doctrine:

The expression "cold" multiply-denotes (a) a mechanical property of snow (and other cold things), and (b) a type of sensory experience. People confuse these denotations, even though only a physical body can have the mechanical property, and only a mental state can instantiate the sensation-type.

Intuitively, an explanation for (1) follows. People are persuaded they can understand what cold is, without reference to the purely spatial and quantitative properties and relations of alleged "natural machine components," because they take cold to be the cold way of feeling. People know the characteristic feel of coldness, and they think that suffices to identify cold. They are convinced that they do not need any information about the micromechanical properties of objects in order to effect this identification.

And, intuitively, an explanation for (2) follows. When people experiment to find out how cold things behave, they learn that they can do so by manipulating physical things such as snowballs. They learn the effect of sticking the snowball in a fire, and so on. It seems right to people to report on these efforts in such terms as "The snowball got less cold real fast in the fire." People are inclined to summarize this with some such slogan as: coldness is objectively in the world; it just seems so from a lifetime's experiences of cold and of the way it behaves.

So mechanism can be reconciled with the appearances, with what common sense "teaches us about cold." But the reconciliation is purchased by representing common sense as confusing a physical property with a way of feeling. Both (1) and (2) are "errors" resulting from the confusion of the common sense idea of cold—which therefore is a paradigm example of a materially false idea. Descartes believed he had found techniques for dispelling the confusions of common sense even while they were still largely in place. The techniques involved a process of gradual scientific self-enlightenment, one of the consequences of which would be a gradual increase in one's ability to discriminate conceptually one term of a confused

pair from the other—for example, physical coldness from the sensation of coldness.

It should seem just a little puzzling how one could manage this process of scientific self-enlightenment, how one could manage to reason one's way to an unconfused worldview, when all one has to work with at the start is a battery of confused concepts. And yet when Descartes presented his epistemology in a sequence of "meditations," in the course of which a naïve meditator is exhibited reasoning his way to enlightenment, the meditator is presented as arguing competently from a strictly logical point of view, making arguments with conclusions that follow neatly from premises. But at the early stages of the self-enlightenment process, the meditator has confused ideas. So it does seem that Descartes believed confusion in one's ideas does not bar validity in one's inferences. The account of confusion I will give later in this book makes room for this combination of conceptual confusion and valid reasoning. Someone who agrees with Descartes that the combination is possible will find this a strength of the analysis I recommend, but someone who is sure that Descartes is wrong, someone who thinks that valid arguments cannot be made with confused ideas, may well reject my analysis, along with the *Meditations.*

Actually, it is hard to find examples of philosophers—even philosophers writing in the early modern period when confusion was a hot topic—who take the strong view that validity is impossible in the presence of confusion. But some did take the view that knowledge is impossible in the presence of confusion, and perhaps soundness in arguments as well. John Locke made it sound as though he held even stronger opinions, for instance when he wrote that "to have [no ideas] distinct, is to have no use of our Faculties, to have no Knowledge at all."[10]

It may be that Locke was resonating to the idea that multiple references, or multiple aboutness, is incompatible with one's beliefs being true or false. Since knowledge requires truth, this would account for the second half of his gloomy opinion. The first half of his opinion, that in the presence of general confusion we would have "no use of our faculties," presumably expresses the notion that if you have hardly any thoughts that qualify as knowings, you do not have a sufficient collection of known premises upon which to build, by inference, a respectable body of further knowledge. Your faculties don't work, really, even if they seem to you to be working. At any rate, these are the charitable readings. An uncharitable reading would take "have no use of our faculties" to mean "be unable to think cogently." If he meant that, then he owed an argument. It is con-

ceivable that he misunderstood Descartes, perhaps imagining that having confused ideas was a matter of walking around as though stunned. There is no evidence to support the uncharitable reading, except his doomsday turn of phrase.

Locke argued at some length that common sense is not confused. In Chapter 19 we will review his argument. Here, let it suffice to say that he had no need of Descartes' maneuver for escaping the clutches of (1) and (2) because he had his own maneuver—a resort to powers. In English-language philosophy, interest in the phenomenon of confusion waned. And never again waxed.[11]

What followed (in English-language philosophy) were several immensely influential empiricist epistemologies. These epistemologies typically elevated simple common sense "perceptual" concepts, such as red or cold, to the status of the most easily and clearly known of all concepts a person can have. To accomplish that elevation, the empiricists took these concepts to be definable entirely in terms of "immediately accessible" sensory content, plus logical apparatus in a broad sense of "logical" that includes, for example, the ideas of tendency or disposition.

To help me streamline the point I want to make, assume that physical science recognizes one and only one physical property by virtue of which red things are red, and assume it is called red*. Then let red# be a neo-Lockean property defined by the formula:

> an object x is red# <=> x is disposed to cause sensations of red in (normal) perceivers (normally).

Assume empiricism holds these two theses:

(1) red = red#.
(2) The meaning of "sensation of red" can be explained without reference to which physical things are red, by paying attention to the character of visual experience itself, some of which happens to be noticeably of-red.

Thesis (2) makes the neo-Lockean formula defining red# noncircular, even given thesis (1).

Since the physical sciences do not traffic in "sensory awarenesses" and other mental items, it is very plausible that red* is an entirely different property from red#, and therefore an entirely different property from red. In that case one can ask how these properties are related. Maybe they are related by some sort of psychophysical law saying that all and only red*

things are red# (and thus plain red). Maybe they are not related by any laws, but nevertheless red# (and plain red) are in some way "reducible to" red*, or "supervene on" red*. Maybe red* is identical with red# (and red) after all; the existence of a red* but nonred# thing, or a red# but nonred* thing, being only an epistemic possibility, not a full-blooded metaphysical possibility.

These metaphysical alternatives are hard to appraise. More than one philosopher has suspected them of being somehow fake. The only reason they can be described is because common sense red is thought of as a quasi-mental property in order to make it a paradigmatically knowable one. Descartes' picture of common sense went the other way: common sense red is a paradigmatically confused concept, not a knowable one at all—if this means "productive of knowledge." The common sense person confuses physical red with the visual sensation of red, just as she confuses physical cold with the sensation of cold. The relation between common sense red and "the red way of sensing" is not that given by theses (1) and (2) and the neo-Lockean definition of red#. Common sense red "is" the red way of sensing, in addition to "being" something else as well. And common sense red "is" red*, in addition to "being" something else as well. But it is not true that the red way of sensing "is" red*, because the relation expressed by "is" in this string of propositions is not transitive. It would be better to put the whole idea in terms of semantic relations, rather than funny, nontransitive "identity," but funny identity suffices to make the main point.

The main point is that nobody would care how the neo-Lockean property red# is related to red* except for thesis (1). The interesting issue is how red, common sense red, is related to red*. Descartes' answer is that the one is the other, so that there is no question of one being correlated with the other, or reducible to the other. Therefore the vexing metaphysical alternatives cannot be intelligibly described.[12]

The price a Cartesian pays to avoid the vexing metaphysical alternatives is epistemological. The clean, compelling empiricist epistemology championed by Locke to replace the scary, confusion-embracing Cartesian epistemology cannot get off the ground unless the vocabulary of color and shape and sound is privileged. According to the empiricist epistemological vision, when you describe what the dials on your laboratory instruments say—for instance when you assert that "The long green pointer is on the red line"—you must be able to rely on your description being largely free

of dubious concepts. It is with such descriptions that your knowledge of the world begins. Descartes won't let you have that.[13]

Nor were the empiricists alone in celebrating the good sense shown by modern philosophy when it chucked Descartes' picture in favor of Locke's. One can almost hear Kant breathing a sigh of relief that the bad old days were gone, the days when "all philosophical insight into the nature of sensuous cognition was spoiled by making the sensibility merely a confused mode of representation, according to which we still know things as they are, but without being able to reduce everything in this our representation to a clear consciousness."[14]

Since Kant, we have concentrated on what might be called propositional epistemology, including the propositional epistemology of perception. In the twentieth century the debate about the "incorrigibility" of sense-content propositions such as "I feel cool" or "I am appeared-to redly" has concerned the epistemic properties of these propositions—for instance, whether they are noninferentially knowable, or whether they are "made true by" the very experiences one takes as confirming them. These were not early modern concerns. Usually, when early modern epistemologists sound as though they are worrying about these issues of the propositional epistemology of perception, they are in fact grappling with the problem of how—if at all—there can be unconfused ("distinct") awarenesses, and the subsidiary problem of how one might figure out which kind one was experiencing. Here, for example, is John Locke: "let any *Idea* be as it will, it can be no other but such as the Mind perceives it to be; and that very perception, sufficiently distinguishes it from all other *Ideas*, which cannot be other, i.e., different, without being perceived to be so."[15]

Identify ideas with sensations and you can get this to say that the truth of such a proposition as "I am appeared-to redly" is constituted by one's sensory experience. But in fact this passage is picking up on a theme harped on at length elsewhere in the *Essay:*

> Every one that has any Knowledge at all, has, as the Foundation of it, various and distinct *Ideas:* And it is the first act of the Mind, (without which, it can never be capable of any Knowledge,) to know every one of its *Ideas* by it self, and distinguish it from others. Every one finds in himself, that he knows the *Ideas* he has; That he knows also, when any one is in his Understanding, and what it is; And that when more than one are there, he knows them distinctly and unconfusedly one

> from another. Which always being so, (it being impossible but that he should perceive what he perceives,) he can never be in doubt when any *Idea* is in his Mind, that it is there, and is that *Idea* it is; and that two distinct Ideas, when they are in his Mind, are there, and are not one and the same *Idea*.[16]

Locke's contention boils down to this: pretend your ideas are the only things in existence, so that if you cannot confuse one idea with another, then you cannot be confused at all. And—glory be—in fact you cannot confuse one idea with another. If you restrict your judgments to the properties of, and relations among, your ideas, then none of these judgments is confused. Ideas are "given" as distinct (unconfused). This supplies you with a big pile of objects for which there is no risk of thinking one thing is another. If you still wind up confused despite having this unconfused substratum of knowledge, shame on you. To help you along, Locke explains how to be sure you have not stepped outside this arena of guaranteed distinctness of objects when you make knowledge claims.[17] The emphasis is on what class or classes of entities can be "distinctly thought," rather than on what class or classes of propositions have certain desirable epistemic properties.

Of course epistemological conclusions about the epistemic properties of propositions may follow as corollaries. Locke draws just such a conclusion here. He argues that the mind's unfailing ability to identify an idea as the idea it is, and to tell one idea from another, is the source of the absolute certainty of propositions which do not go beyond affirming that some idea is indeed itself, or that it is not some other idea. Some examples he cites are: *whatsoever is white is white, a man is not a horse, red is not blew.*[18]

II

WHAT CONFUSION IS

3

Fred and the Ant Colony

It will be helpful to set up a case of confusion that is simple enough to expedite our intuitions about how various features of the case contribute to the plausibility of an attribution of confusion, but realistic enough that few important aspects of confusion are missing. My favorite toy case for these purposes is the story of Fred and the ant colony.

A guy named Fred sees an advertisement for an ant colony in the back of a magazine he is reading in the dentist's office. He sends away for it, and when it arrives he sets it up in his kitchen. The ant colony consists of a glass box that Fred fills with dirt and some leaves and twigs from his back yard, a bottle cap for water, and some ant food from a little envelope that came with the kit. The ants arrived in two separate containers, one labeled "little" and one labeled "big." Never reading the labels, Fred dumps the contents of both containers into the ant colony. He notices that there is a large number of little ants, as one would expect of little ants, and—he believes—one big ant, much bigger than the others. But Fred was inattentive and missed a second big ant when it tumbled into the ant colony.

Everyone knows that big ants make every effort to avoid conflict. Fred's two big ants immediately arrange to split their time between running around on the surface performing feats of strength, and napping down in the bowels of the ant colony. They alternate up-time, one down below whenever the other is on the surface. As a result, even after Fred becomes fascinated by "the" big ant, studying "its" movements as often as he can, he never catches on that there are in fact two big ants. He decides to name "the" big ant "Charley." It is best to imagine Fred introducing the name "Charley" while he is off somewhere out of sight of his ant colony; then there is no temptation to think of Fred as using "Charley" as the name of whichever big ant is before his eyes when he first employs the name. We might even add that Fred is not picturing either of the big ants when he

first uses the name "Charley," nor has he just deployed a description which in point of fact fits only one of them.

So much for Fred. We, being more astute and more ant-savvy, watch Fred set up his ant colony and notice the two big ants go in. We even notice that one of them is a tad skinnier than the other, although only someone looking for the difference, and as keen a judge of ants as we, would ever be able to tell. We decide to call the big ants "Ant A" and "Ant B." But we decide against clueing Fred in. Although it does not put us in a very good light to say it, we rather enjoy making fun of Fred's inability to tell that Ant A is not Ant B, his carrying on about Charley, his theorizing about the social relations between Charley and the many smaller, weaker ants. We find it especially amusing when Fred looks at Ant A (Ant B) and says "Charley, there, did X earlier today," or something along those lines, when—as we know—it was Ant B (Ant A) who actually did X. Fred, we say, thinks Ant A is Ant B. He is wrong to do so, we believe—not that we can say very much about the kind of "wrongness" this is, except that it is a matter of confusing things. Let this be our working example of a confusion, and let our judgment "Fred thinks Ant A is Ant B" be our working example of an attribution of confusion.[1]

An aside on method. "We" are playing two different roles here. We—you the reader and I the author—are characters in my fictional story about Fred and his ants. As characters in that story, we make certain assertions and hold certain beliefs. In particular, we say, and believe, of Fred that he "thinks Ant A is Ant B." But we—again you the reader and I the author—are, of course, the reader and author, respectively, of my fictional story about Fred and his ants. As such, we will make (I will urge you to make and you, I hope, will go along) certain judgments about things said and believed by characters in the story, such as the judgment that a given statement is natural and correct for "us," as characters in the story, to make. When we, as reader and author, make such judgments, we may sometimes know certain things about the world of my story that no character in the story knows, as the story is being told. When one is doing epistemology, it is important to keep this possibility in mind, so that we, as philosophers reflecting upon a fictional story as a philosophical technique, do not confuse "we" the characters with "we" the philosophers, thereby unwittingly misconstruing what would or would not be rational for "us" the characters to think or say. If a serious risk of this error arises, I will flag it.

One could tell a different story; something along these lines: for a long time Fred has an ant colony containing only one big ant. Everyone, Fred

included, calls this big ant "Brewster." Next door a neighbor has an ant colony containing one big ant that everyone, Fred included, calls "Dexter." Dexter escapes, goes next door, and sneaks into Fred's ant colony. Fred sees Dexter in the morning, misidentifies Dexter as Brewster and does not catch his error for some time, Brewster having run down into the bowels of Fred's ant colony. The rest of us are fully aware of these events (somehow), so we are amused when Fred says such things as "Why Brewster, you're eating as though you spent the night running around"—all the while looking squarely at Dexter.

This is essentially the setup Saul Kripke uses to introduce a distinction between semantic reference and speaker's reference. The reference of the name "Brewster" in common parlance is Fred's big ant, so we can say that the name "Brewster" semantically refers to that ant, no matter who uses it, Fred included.[2] Fred is subject to the norms of the common language regardless of what errors he may have made; it's his language. The same goes for Fred's use of the name "Dexter." But Fred uses the name "Brewster" with a speaker's reference that does not coincide with the reference the name has in the common language. He uses it to refer to Dexter. He did not use it that way before today, and he may not use it that way much longer, but for now he does. Fred has acquired a short-term referential practice of misapplying the name "Brewster"—or rather, it is a misapplication, provided the standard for correct application is taken to be fixed by the long-term referential practice Fred shares with the rest of us.

The Brewster/Dexter story could well serve as our working example of confusion. All the same points can be made, I believe, but with an unnecessary additional level of complexity. Versions of options (1)–(5) of Chapter 1 can be described just in terms of the concept of "speaker's reference."

The easiest way to generate speaker's-reference versions of options (1) through (5) is to focus on one particular purpose we might have for assigning a speaker's reference to Fred's use of the name "Brewster." Assume that we are especially concerned to explain why Fred makes perceptual reports using the name "Brewster" when in fact it is Dexter who is exhibiting the behavior described in the report, and indeed Dexter is doing so before Fred's eyes. We will do well to pick as "the referent" of Fred's use of "Brewster" whatever object X best satisfies this criterion.

The predicative content of those of Fred's perceptual beliefs he expresses using the term "Brewster" covaries with the properties Fred perceives X to have, other things being equal.

The effect of this criterion is to guarantee (or almost guarantee) that Fred's "Brewster" thoughts and "Brewster" statements mirror his (Fred's) observations of the object X. This fairly weak connection between thought or talk on the one hand, and perception of objects on the other, seems a plausible constraint to place on any explanation of perception-driven cognition.

We should have a special label for the kind of reference our criterion addresses, since it might differ from the "semantic reference" of a name. "Speaker's reference" is as good a label as any. Now, the object X that best satisfies our criterion is Dexter. So we ought to cite Dexter as the speaker's referent of Fred's use of Brewster.

There is no reason why we shouldn't attach great importance to giving explanations of Fred's perceptual beliefs concerning the overt behavior of ants running around in front of him. Indeed, there is no reason why we shouldn't care about it exclusively, with nary a moment's concern about other matters that might have motivated us to make other, different assignments of speaker's reference. But suppose we do have other, equally weighty concerns.

For instance, suppose Fred is downtown shopping a few hours after first seeing Dexter in the ant colony. Fred says "Brewster is a very strong ant." If pressed, he will cite various past observations of the authentic Brewster performing feats of strength, as well as various past observations of Dexter. Suppose we want very much to understand why Fred has said what he just said. But although Fred's beliefs are perception-based, the perceptions lie in the past. So our criterion does not apply. We must revise our criterion slightly, putting the verb "perceives" in the past tense. But it is at least possible that the belief Fred is reporting arises from a combination of Brewster-experiences and Dexter-experiences in roughly equal measure. Then our criterion does not pick out any one object X as the speaker's referent of Fred's use of the name "Brewster"; Dexter and Brewster are in a tie. We could look for some new tie-breaking criterion which would enable us to assign one or the other as the speaker's referent. Or we could assign both, letting Fred's present use of "Brewster" have a multiple speaker's reference. Or we could assign neither, and do without the concept of reference in our explanation of object-perception.

Plainly we have just replicated the spectrum of options (1)–(5) of Chapter 1. But we have framed them as options for assigning a speaker's reference, declining to deal with the idea of semantic reference, or reference in the common language. As I said, this seems to me an unnecessary compli-

cation. So in the story of Fred and his ant colony as I actually have told it, Fred mixes up his two big ants from the start and coins a name, "Charley," for what he supposes to be "the single big ant." This is all the meaning the name "Charley" has in the common language. We get to ask our philosophical questions just the same, but we do not have a second dimension of reference or meaning floating around needing to be watched. Now back to the thread.

Fred's confusion is not a matter of his holding the false belief that there is exactly one big ant in the ant colony.[3] Argument: Consider a revised story in which Fred never so much as glances at either of the big ants. But he believes that there is exactly one big ant in the ant colony because we tell him it is true and he believes everything we say. (I am making this farfetched just to make it quick.) In the circumstances of this revised story, Fred would not have confused Ant A with Ant B. Therefore accepting the false numerical proposition about the big ants is not the same state as confusing them, since the former state can occur in the absence of the latter.

Nor is Fred's confusing Ant A with Ant B a matter of his holding some combination of false identity beliefs, such as "The ant that ate the raisin [this was A] is the ant that drank the water [this was B]; and the ant that lugged the stick around [this was A] is the ant that lugged the pebble around [this was B]; and so on." Argument: Let this be the proposition P.

(P) The ant that ate the raisin is the ant that drank the water; and the ant that lugged the stick around is the ant that lugged the pebble around; and (etc.).

Revise the story so that (a) Ant A and Ant B do not exist; (b) a whole slew of diverse ants did the various things enumerated, rendering proposition P false; (c) Fred accepts P (which is a lie) because he believes the lie. As an alternative, you can let Ant A and Ant B exist, but stipulate that Fred has never seen either one, and stipulate further that neither Ant A nor Ant B has anything to do with the truth or falsehood of P. In either of these revised stories Fred would not have "confused Ant A with Ant B." So, merely having the false belief that P would not suffice to qualify Fred as having confused Ant A with Ant B.

These arguments generalize easily. Take any set of false propositions. In a revised story Fred might accept these propositions for some reason having nothing to do with confusing Ant A with Ant B; indeed, in circumstances where he had never seen either of the big ants, or where he had seen only one of the big ants and thus never confused it with the other.

Two points need to be made here, to guard against possible misunderstandings:

First point. The form of each of the arguments we just surveyed is: "Fred could be in belief-state of type B when he is not confused; so being confused does not consist in being in belief-state of type B." As Saul Kripke reminds us, the argument form "State-type X could be exemplified when state-type Y is not; therefore these state-types are distinct from each other" is invalid when the "could" expresses merely epistemic possibility.[4] But in our arguments the "could" does not express merely epistemic possibility. The premise is not "it is compatible with everything we know that Fred be in belief-state of type B when he is not confused" (although that epistemic claim is true). The premise is "being in belief-state of type B (positively) does not suffice for being confused."

Second point. Some philosophers want to speak of "contingent identity," to be contrasted with "necessary identity." For example, someone might want to be able to say that the computational state CS exhibited by a certain machine at a certain moment is *contingently identical to* the physical state PS of the machine hardware which "realizes" the computational state. The philosophical motivation might be to underwrite a metaphysical claim to the effect that there is nothing there but physical stuff, without being committed to the unacceptable thesis that one could adequately *define* or *conceptualize* the computational state by describing the physical state in the terms of, say, materials engineering.

One should not argue from the premises "this object could exhibit state-type S_1 when it does not exhibit state-type S_2" and "this object exhibits state-type S_1" to the conclusion "state-type S_1, as exhibited by this object, is not contingently identical to state-type S_2." One should not argue that way even when the sense of "could" in the first premise is "logical" or "metaphysical" (that is, more than merely epistemic). The example of computational states and their physical or material "realizations" is enough to show this. Our arguments aimed at showing that Fred's confusion does not consist in his holding a false propositional belief had better not run afoul of this restriction on reasoning involving contingent identity. And they do not. The point of our arguments *precisely is* to show that "having such-and-such type of belief" does not adequately define or conceptualize being confused. That being our purpose, we are not interested in learning what confusion is or is not contingently identical to.[5]

So confusing one thing with another is not a matter of having false be-

liefs "de dicto"; it is not a matter of accepting false propositions. Or, to speak more accurately, an attribution of confusion does not attribute a state that can be characterized as a belief in a set of false propositions (either separately or conjoined—though I did not distinguish those cases in sketching the arguments).

It is traditional to contrast attributions of belief *de dicto* with attributions of belief "de re," whereby the believer is characterized as believing of some object that it has a certain property, or is characterized as believing of several objects that in a given order they satisfy some relation. Is the remark "Fred thinks Ant A is Ant B" a *de re* belief attribution?[6] If it is, then presumably it says Fred believes of Ant A and of Ant B that they satisfy the identity relation. Does that make sense?

It is impossible to get a grip on the concept of *de re* belief without a prior grip on the concept of the reference of the believer's terms. To see why that matters here, recall the options we surveyed in Chapter 1 for assigning reference and truth value in the presence of confusion. In particular, recall options (1), (2), and (3). We can set up an argument in the form of a dilemma. First we assume that either option (1) or option (2) is right—they are interchangeable here—and on this assumption we show that confusion should not be identified with *de re* belief. Next we assume that option (3) is right, and on this assumption, too, we show that confusion should not be identified with *de re* belief. The first horn of the dilemma is a snap. The second is not, but we'll have a try at it.

First horn: Fred uses "Charley" to refer to Ant A and only Ant A. And Fred uses such expressions as "that big ant right there" to refer to Ant A always, regardless which big ant he is pointing at. (The assumption that he uses these expressions to refer to Ant B and only Ant B plays out the same way, ants reversed.) Then how does Fred manage to think a thought "of Ant B"? If he cannot, and it seems he cannot, then he cannot think of Ant A and of Ant B that the former is the latter. Indeed he cannot think anything at all, either monadic or relational, of Ant B.

Second horn: suppose Fred uses "Charley" and his other big-ant expressions to refer multiply to Ant A and to Ant B at once (again regardless of which, if any, big ant is before his eyes). One intuition suggests that since Fred cannot speak of Ant A *and just it,* or of Ant B *and just it,* he cannot have a thought of Ant A *and just it,* or of Ant B *and just it.* And since he cannot "single out in thought" either one of the big ants, he cannot think, with respect to either one of them, that it "taken by itself" has any prop-

erty at all, including being identical with the other big ant. So he cannot think such a thing falsely. This set of intuitions leads to the conclusion that Fred's confusion ought not be identified with "false *de re* belief."[7]

This way of arguing the second horn of our dilemma relies on the proposition that a person who has a belief with respect to a given object, to the effect that it has some property, must have that belief with respect to the given object "and just it." But is this really a necessary condition on "belief with respect to"? It is easy to imagine Descartes denying that it is, and insisting on a wider use of "with respect to" that would accommodate the mental acts comprising common sense, despite its universal confusion. Indeed I described Descartes' opinion in very much those terms in Chapter 2.

Of course, even if we decide that "belief with respect to" is compatible with multiple reference, we still would not have decided whether such beliefs ever have truth values. So we would not have decided whether they can be false. Option (3) of Chapter 1 included the suggestion that we couple the idea of multiple reference with the idea of assigning truth values by supervaluation. Let's stick with that suggestion. That makes it easy to evaluate a sentence like "Charley is Charley" as Fred uses it: the sentence is true, since Ant A is Ant A and Ant B is Ant B. And it makes it easy to evaluate a sentence like "Charley ate a raisin earlier today," when Ant A did but Ant B did not. The sentence is neither true nor false. How should we apply this line of thought to help us assign truth values to beliefs attributed "de re"? For example, suppose we want to consider the truth or falsehood of the "identity" belief one attributes to Fred by saying, in the *de re* tone of voice, "Fred believes that the ant that ate the leaf at two o'clock is the same ant as the ant that drank from the bottlecap at six o'clock." When is this belief true, when is it false, and when is it neither? Let's explore a bit.

Here is a view to consider, so as to reject it: when someone says "Fred believes that the ant that ate the leaf at two o'clock is the same ant as the ant that drank from the bottlecap at six o'clock," and this is a *de re* attribution, the speaker is using the definite descriptions "the ant that ate the leaf at two o'clock" and "the ant that drank from the bottlecap at six o'clock" to refer to exactly the objects she, the speaker, takes these descriptions to fit. This is a "normal" way to think about *de re* attribution of belief. Belief attributions *de re* usually are made with respect to things of which the belief-attributer means to speak, and therefore things the attributer distin-

guishes in her own thought and separately denotes in her own language. But if we are going to talk about *de re* attributions of belief in the presence of confusion, some changes must be made in that standard picture. In the present example, the problem is that if the belief-attributer is using the definite descriptions to refer to items she can distinguish (Ant A and Ant B), then her belief statement represents Fred as having beliefs that discriminate between Ant A and Ant B. This is not in the spirit of holding Fred to be confused. We need a different way to look at what the belief-attributer is doing.

A reasonable suggestion is that the belief-attributer is using the definite descriptions to refer multiply to both big ants, doing so on behalf of Fred, as it were, since she wishes to report a mental act of which Fred is cognitively capable, even though she is making her report *de re*—"of things"—in some sense. So let's treat the description "the ant that ate the leaf at two o'clock" as a multiply-denoting term denoting both Ant A and Ant B; and ditto for "the ant that drank from the bottlecap at six o'clock." Let's treat the descriptions this way even if (say) Fred watched Ant A, and only Ant A, eat the leaf. Finally, let's assign truth values to Fred's beliefs, when they are attributed *de re,* by assigning whatever truth value supervaluation tells us we should assign to the subordinate clause of the belief sentence, with the proviso that definite descriptions one would ordinarily assume fit one or the other of the big ants in fact multiply-denote both big ants at once. That is a windy formula, and limited in scope, but good enough for our limited purposes here.

Now suppose Ant A ate the leaf but never drank from the bottlecap, whereas Ant B drank from the bottlecap but never ate the leaf. Our rule for assigning truth values to Fred's beliefs tells us that his belief that "the ant that ate the leaf at two o'clock is the same ant as the ant that drank from the bottlecap at six o'clock" is neither true nor false. Therefore it is not false. But the belief in question surely is paradigmatic of the *de re* "identity" beliefs that might, we thought, all be false, and taken together might add up to, might constitute, Fred's being confused. That idea is wrong. Or rather, it is wrong so far as one can tell, bearing in mind the difficulty of applying the concept of a *de re* belief in the presence of confusion.

But even though these identity beliefs of Fred's are not false, they are untrue (and unfalse). So maybe we should equate Fred's confusion with his holding many untrue and unfalse *de re* identity beliefs "about" Ant A

and Ant B. The problem now is that this fact—the fact that Fred holds all these untrue and unfalse identity beliefs—cries out for an explanation. You and I are not like that; why is Fred? Here is a very persuasive explanation.

Fred thinks Ant A is Ant B; he has the two ants confused. So whenever he obtains information about the doings of one big ant (perhaps by watching it), he takes this for information about the doings of the other big ant as well. When we report Fred's beliefs we (may) encode this aberrant uptake of information by making a *de re* attribution of an identity belief, something of the form "Fred believes that the F ant is the G ant," where Ant A (Ant B) is F but Ant B (Ant A) is G. The truth value of the belief we attribute must be computed supervaluationally on the assumption that the expressions "the F ant" and "the G ant" multiply denote. Evidently, then, these beliefs we attribute will lack truth value. Nothing similar befalls you or me because we are not similarly confused.

If Fred's confusion just *is* his holding these untrue and unfalse *de re* identity beliefs, this explanation is circular, since it invokes his confusion to explain why he holds the beliefs. But it is not circular. On its face it is very plausible. (Though only on its face. See note 8 below.) This completes the second horn of our dilemma.

In sum: the suggestion that attributions of confusion are *de re* attributions of false (or maybe untrue and unfalse) identity belief is not promising. Since the suggestion that attributions of confusion are *de dicto* attributions of false belief also is unpromising, we ought to conclude that attributions of confusion are not attributions of false or untrue belief at all. Confusion is not false belief. It follows that the "error" of which a confused person is guilty is not the error of believing the false. It follows further that correcting a person's confusion is not a matter of leading the person to substitute true for false beliefs.[8] Standard epistemology loses most of its tool kit in one fell swoop. We need to start building a new one.

4

The Semantic Use of Psychological Language

When we (we in the story) attribute confusion to Fred about the ants, we are not attributing a psychological propositional attitude, such as believing a false uniqueness or identity proposition, or a string of them. That is surprising, because the sentence "Fred thinks Ant A is Ant B" seems to ascribe a state of mind. But this is not such an unusual case. The subordinate clause following "thinks that" (and many other "psychological" verbs) often is best thought of as supplying a type of information about the subject that has nothing to do with what propositions the subject accepts, or doubts, or knows, or wishes were true, or fears are true. Nor does the subordinate clause supply information "de re" about what entities the subject bears a mental attitude toward. In the cases I have in mind, it is natural to say that psychological language is not being used to say anything "mental" at all—although there is no disadvantage, except surface paradoxicality, in saying that these cases show that sometimes a person's mental state has nothing to do with *how* her mind works. Here is an example, where the operative verb is "thinks," just as in "Fred thinks Ant A is Ant B":

A young child competently performs arithmetic computations with both positive and negative integers. She is good at division, and when, as she says, one number "can't be divided" by some other number, she dutifully expresses the answer with a remainder. We ask her whether every number can be divided by two. "No," she answers.

We say: "She thinks all numbers are integers." What does this mean? Obviously it does not mean that she accepts the proposition "All numbers are integers," assuming that she has never heard the word "integer." But maybe she has heard the words "whole number." Assume she has. But assume also that she has never seen a fraction, or been taught anything at all about fractions under any description.[1] She does not know about any kind of number that contrasts with a "whole number," so the only meaning she

can attach to the sentence "All numbers are whole numbers" is "All numbers are numbers," or "Numbers are what they are," or something similar. This is not the meaning we attach to the sentence "All numbers are integers (or "whole numbers"), so the child does not and cannot comprehend the proposition we express by that sentence.

So, again, what do we mean when we say "She thinks all numbers are integers"? Is there some other belief concerning numbers, or more generally concerning arithmetic, we mean to be ascribing to the child? Or some set of such beliefs? I will not take the time to survey possible candidates. I cannot think of any that survive even cursory examination. The reason is that when we say of the child "She thinks all numbers are integers," we are performing a speech-act that is not a mental-state ascription at all. We are doing what I will call "taking a semantic position."

A semantic position is a position, a "stand," on how someone's reasoning ought to be appraised or what criterion of argument-goodness ought to be applied. Since our concern is only with the child's arithmetic reasoning and not with her reasoning on other topics, the species of argument-goodness at issue is validity. When we say that the child thinks all numbers are integers, we are taking the position that her arithmetic reasoning be evaluated for validity by means of a criterion more forgiving than the criterion appropriate for the arithmetic reasoning of an educated adult. In particular, we are proposing a domain restriction: the domain is to be restricted to integers. Throw out the other numbers, such as rationals not equivalent to an integer, and restrict the arithmetic functions to the integral subset. Other alterations of the normal adult validity criterion would be motivated by this one.

To take this semantic position is to adopt an attitude of inferential charity toward the child. We must contrast inferential charity with doxastic charity. To be doxastically charitable is to insist that a subject's language be interpreted so as to maximize the truth of her beliefs, or the truth of some select, "really important" subset of her beliefs. To be inferentially charitable is to insist that a subject's language be interpreted so as to render valid (or, perhaps, nonmonotonically strong) certain classes of argument which would be rendered invalid (or nonmonotonically weak) by a "standard" interpretation. Doxastic charity plumps for counting the subject as right; inferential charity plumps for counting her as reasonable. We will see that the presence of confusion in a subject typically compels one to choose between these modes of charity. You can make them right, but then you can-

not make them reasonable; or you can make them reasonable, but then you cannot make them right. But that is getting ahead of ourselves.

The child draws the inference "The answer is three, so the answer can't be divided by two." Good for her! In fact, good for her every time she thinks like that. The question is, thinks like what? A full answer must take the form: whenever she gives an argument with such-and-such semantic feature, to be called "validity"—validity measured by a charitable criterion. The remark "The child thinks all numbers are integers" contains precious little detailed information about how a suitably inferentially charitable semantics might be drawn up. That is because it is brief, and because it is a nonlogician's remark. But it is not empty: it has content enough for a logician to go ahead and fill in details, perhaps by first giving a semantics appropriate to the arithmetic reasoning of a normal educated adult, then detailing how to restrict the domain as recommended without messing up any of the other semantic machinery. The nonlogician who says "The child thinks all numbers are integers" can point to the details the logician provides and say, "Right; I'm with her." More about this deferential piggybacking in a moment.

When we take the position that the child's inferences be evaluated for validity in a charitable fashion, we are demeaning the child as well as demanding fair treatment for her. It is clear from the circumstances that we take inferential charity to be needed because we think (a) the child is not up to the task of satisfying a full-fledged grownup validity criterion, and (b) this semantic infirmity should not be held against her, she does not merit a barrage of "that just doesn't follow" criticisms. To think (b) is to think well of her, in a sense, and to stand behind it. But to think (a) is to denigrate her rationality, to hold her out as rationally defective. In a word, we are being paternalistic.

Inferential charity is not always paternalistic. Suppose a mathematician begins her lecture with the remark "All groups are commutative."[2] This is a stipulation of a domain restriction, and therefore a plea for inferential charity. But the mathematician does not mean to denigrate her own rational faculties, or those of what we may imagine is her learned audience. She just doesn't want people testing the arguments she gives today against a universe of discourse with a bunch of noncommutative groups floating around in it.

When the verb "to think" or some other psychological verb is used to take a semantic position, I will say that it has its "semantic use" rather than

its "psychological use." An attribution of belief, even an attribution of a cluster of beliefs, is a *local* characterization of the subject's mental state: it does not concern the subject's language or thought taken as a whole. By contrast, someone who takes a semantic position recommends a *global* characterization of the subject's language, and therefore of the subject's mental states. A validity criterion applies to any (verbal) reasoning the subject does, and therefore has implications for the reasonableness or unreasonableness of any belief the subject has. This is why one cannot put one's finger on any particular beliefs, or even any particular set of beliefs, which are "the very beliefs" one is ascribing to the child when one says that she "thinks all numbers are integers."

I suspect a failure to distinguish the semantic use of psychological language from the psychological use has led some philosophers to seize upon what they rightly perceive as a "globally characterizing" element that is present when psychological verbs are doing semantic work. They then rashly infer that some sort of holism is true of psychological language when it has its psychological use—a holism according to which every mental-state ascription, or perhaps just every intentional-state ascription, has commitments that pertain to the whole of the subject's psychological state. But that is a long, sad story, and is best left for another occasion.

When we (in the story) say of Fred that he "thinks Ant A is Ant B," we are using the verb "thinks" semantically, not psychologically. We are taking the position that Fred's reasoning concerning his ant colony should be appraised for argument-strength by means of a special charitable criterion. But here the idea is *not* that a domain restriction is in order. The idea is that Fred's arguments should be evaluated for inferential strength by a validity criterion with these general properties (whatever the details may be):

(a) The criterion represents Fred's "Charley" inferences as pertaining to Ant A and Ant B at once; and as a result,
(b) The criterion evaluates Fred's "Charley" inferences as valid exactly when (or almost exactly when) they are of a form we would regard as valid in the absence of any confusion, where "Charley" was just an ordinary name.

Here's a gloss:

First point. I mean the language "pertaining to Ant A and Ant B at once" in (a) to have a broader meaning than "denoting both Ant A and Ant B at once." One way a semantics for Fred's "Charley"-talk might represent Fred's inferences as pertaining to both ants at once would be to

characterize some expressions, especially the name "Charley," as multiply denoting. But there are other ways, and some make no use of the denotation relation at all. I will recommend just such a semantics in Part Five. The prephilosophical comment, "Fred thinks Ant A is Ant B," surely is not meant to take sides in a philosophical dispute between a philosopher with my view of the matter, and other philosophers who believe that some of Fred's terms should be understood as multiply denoting. Hence the broad language "pertaining to Ant A and Ant B at once."

Second point: clause (b) is a way of formulating the requirement of inferential charity. For contrast, it is helpful to consider the effect of an uncharitable validity criterion for Fred's "Charley" inferences. Suppose we were to characterize Fred as equivocating: using the name "Charley" with studied ambiguity. For simplicity, suppose we care only about those "Charley" sentences Fred tokens when he is looking squarely at one or the other of the two big ants. We adopt the policy of interpreting his tokens of the name "Charley" as referring to the ant before his eyes, and only that ant. We accompany this policy with a straightforward classical first-order validity criterion (all that matters to the example is that equivocation be fallacious). Now suppose Fred utters a token of the simplest imaginable inference:

Charley is eating a raisin
Therefore
Charley is eating a raisin.

Fred is a careless observer and thinker, though; even as he speaks the big ant he has been observing scoots below the surface, its colleague scoots up, and immediately finishes the first big ant's meal of tasty raisin. The exchange of ants occurs precisely as Fred says "Therefore." Fred has committed a fallacy of equivocation. Our validity criterion takes him to task for it, and counts his argument as invalid.[3]

The idea behind clause (b) is this: to say that Fred "thinks Ant A is Ant B," or "doesn't know the difference between Ant A and Ant B" is, in part, to characterize Fred as unable to make a mistake of logic of the kind we just imagined attributing to him (and to characterize him as abnormal, maybe subnormal, in consequence), just as to say the child thinks all numbers are integers is to characterize her as unable to make certain logical mistakes, because they are mistakes only if one's domain of numbers is more inclusive than the class of integers. In the case of Fred, and in the case of the child, the words "unable to make," as I just used them,

have a semantic—a "logical"—meaning, not a "mentalistic" meaning. It is not that Fred lacks the brainpower to equivocate between two senses of "Charley"; it is that we propose to interpret his language in such a way that the possibility of such an equivocation does not exist. Let the two ants shuffle themselves as they please, we urge; if Fred gives the argument just displayed, it should be counted valid, just as though the name "Charley" denoted a single object. *How* a criterion of validity could be designed so as to get that effect is another matter, and a job for the logician.

Bumper sticker: confusion is a defect in one's "powers of discrimination" which entails no defect of pure reason. This bumper sticker is misleading because it suggests that attributing confusion to a person is commenting on the person's health, like calling the person nearsighted. Other than that, it's a good bumper sticker.

Actually, a person who attributes confusion to Fred, or to anyone, is taking a broader semantic position than I have just described. There is more to being inferentially rational than giving valid arguments. Many arguments are intended only as nonmonotonically strong; that is, the premises are intended to be weighty considerations in favor of the conclusion, with the understanding that this weight can be overridden by other considerations that might pop up. Therefore the scope of our inferential charity toward Fred must be broader than is suggested by (b) above. Our position must be that despite his confusion of Ant A with Ant B, (i) Fred can reason completely logically, and, more broadly, (ii) he can make completely reasonable assessments of the weight of evidence (though perhaps he cannot make completely reasonable assessments of the truth, the credibility, or the accuracy of evidence). The language "assessment of the weight of evidence" is meant to be a quick way of referring to nuanced conclusory talk—such talk as:

"Blah, blah, blah; *and so, surely,* bluh,"
or
"Blah, blah, blah; *and so, likely,* bluh,"
or
"Blah, blah, blah; *and so, maybe,* bluh."

Fred, we maintain, can get this nuanced conclusory talk right, just as right as we can.

Except for the argument of Chapter 7, I will stick to the narrower formulation (b). I will pretend that inferential rationality is a matter of reasoning validly, period. I have no choice, since I do not know how to flesh

out the details of a semantics for nonmonotonic reasoning. Obviously I run a risk. A given semantics might yield a very reasonable account of monotonic validity, but not generalize smoothly to an account of nonmonotonic strength. There's no helping it.

To sum this up: when we say of Fred that he thinks Ant A is Ant B, we take an evaluative attitude toward him, though the type of evaluation we intend is semantic. Similarly, when we say of the child that she thinks all numbers are integers, we take a semantic-evaluative attitude toward her. In both cases there is a quick, formulaic way to voice our evaluative attitude, using the verb "thinks" semantically. In both cases the attitude is paternalistic. We urge that Fred's confusion not be allowed to undermine the validity, the inferential strength, of his reasoning. But in urging this special treatment, we make it clear that we regard Fred as abnormal, needing to be handled with semantic kid gloves. This is the customary paradox of paternalism: the subject is treated as defective *by being accorded* a right to escape negative criticism not accorded to big people. That is the kind of "error" confusion is. It is an error the way childhood is an error. That would please Descartes.

If the title of this Part is read in the most natural way, it is false advertising. I am not trying to explain what person-state is attributed by attributions of confusion. No person-state is attributed. An attributer of confusion is taking a position on how someone, the allegedly confused subject, ought to be semantically characterized; the leading idea of this semantic position is that the subject should be interpreted as "in principle" fully logical, fully inferentially rational (the meaning of "in principle" being that the distinctness of the objects the subject is said to be confusing should not be held to invalidate reasoning that otherwise is of a valid form). So what I am offering is an account of what an attributer of confusion is doing, but I am not offering an account of what an attributer of confusion is attributing.

Here is an obvious objection: the person who attributes confusion to Fred (for instance, "we" the story characters), or the person who says of the child that she thinks all numbers are integers, may well be innocent of logic. How can such a person "take a semantic position," when "a semantic position" is understood to be a position on how the subject's reasoning ought to be evaluated logically? Surely someone could tell that Fred thinks Ant A is Ant B but still have not the slightest notion what a "valid" argument is, as logicians use the term "valid."

The correct response to this objection is that semantic positions are a

subclass of a more general class, the class of political positions, in a broad sense of the term. The form of a political position is, "Let's treat these people this way." A person who takes a semantic position is saying, in effect, "Let's treat this person (or sometimes "these people") this way when it comes to evaluating the goodness of her (their) reasoning." Political-position statements are not very sensitive to the precise wording chosen, because they belong to the political process. You can merely say "I'm against welfare" while someone else issues a lengthy White Paper opposing the welfare state; the two of you may well be taking the same position. Political positions are not individuated like propositions, even though the phrase "what so-and-so means" frequently refers to the political position so-and-so is voicing.

The two paradigm types of political position-taking are: (1) the issuing of a policy statement by some individual who is acting as spokesperson for a group, and (2) a statement made by an individual who means to align herself with some group on a given matter of political policy. Political positions are best thought of as the "content" of certain "expressive activity" on the part of groups, rather than as the content of the expressive activity of some individual. (The plural in the formula "Let *us* treat these people this way" is not accidental.) A political position is taken up by *individuals* in a derivative sense, typically via one of the linkages (1) or (2) just mentioned. Therefore in paradigm cases there will be a great many minds and a great many mouths available to elaborate and defend any given political position. One person—a person who serves as an officer of some antiwar organization—may issue a press release, while another person marches with a crowd chanting "Hell, no, we won't go." But these two people are for the same thing, the position of one is exactly the position of the other, and their opponents will certainly know it. Even if the press release makes reference to concepts the marcher-chanter does not possess, that press release is an articulation of the marcher-chanter's very position.

Sometimes a single individual takes a political position nobody else has ever taken, and makes a statement of it as a call to arms. That is what one does when one "informs" an audience that a person is confused. One calls upon the audience, and any other interested parties, to join with one in applying certain nonstandard criteria when appraising the reasons the subject gives for things she says or believes. If one does not know exactly what this business of "appraising reasons" comes to, or rather, if one does not know this in the detail and depth expected by professional philosophers, no matter. A philosopher can join up and draft the fancy press release. Nor

does it take a special understanding between speaker and audience to secure that right for the philosopher when she finally comes along. Every person with a little experience in politics—which is to say every person not raised by bees—knows that to get others to join in some political position, it is essential that those others be free to express the position in their own terms and defend it in their own voices. If you hold a political position, you are obliged, by the nature of the activity, to allow others with whom you are or come to be aligned to make a great variety of alterations and innovations in what you imagined the position to "be." The point of politics is unity, not pride of authorship.

There is more than one thing to mean by "what so-and-so meant." One might mean "the proposition so-and-so expressed by her words." Or one might mean "the position so-and-so took by uttering her words." Propositions expressed and positions taken are individuated very differently. So, for instance, imagine that when we (we in the story) say of Fred "he thinks Ant A is Ant B," we are entirely ignorant of logical theory. We never have heard of "validity" in the logician's sense of that term; it never occurs to us to give a sustained critique of someone's reasoning as opposed to the person's opinions; perhaps we would not recognize such a critique for what it was if we heard it. It would be foolish to suggest that we (in the story) use the sentence "Fred thinks Ant A is Ant B" to *express the proposition that* Fred's reasoning should be charitably evaluated for validity, by means of a validity criterion which is blind to the distinction between Ant A and Ant B. But it is not foolish to suggest that we (in the story) use that sentence to *take the semantic position that* Fred's reasoning be charitably evaluated for validity, by means of a validity criterion which is blind to the distinction between Ant A and Ant B. Others who are better able to set forth the details of that semantic position—that *very* semantic position—can do so on our behalf. That is exactly what we (author and reader, not we in the story) are about to attempt.

We, the philosophers, will try to supply the semantic details that we, the confusion-attributers in the story, have not supplied, and cannot be expected to supply. A sharing of theoretical labor of this sort is to be expected when a (broadly) political proposal has been put on the table. I will make some concrete suggestions about how the semantic details should be filled in. But my argument will be incomplete, and in any case other philosophers might prefer to supply rather different semantic details. To the extent that the "high theorists" dispute among themselves what theory best fits the political position they are out to articulate, the position taken

by the ordinary, untheoretical speaker—in this case the ordinary, untheoretical confusion-attributer—is underdetermined. Probably that is the best way to look at the meaning of the remark "Fred thinks Ant A is Ant B." If there comes to be settled expert opinion on the question how to describe the details of the logic one is recommending when one utters those words, then we will know exactly what the words mean. At present there is no such settled expert opinion, so the meaning of the words—the position taken by uttering them—is underdetermined; but only in the sense that the logic one recommends applying to Fred's inferences is underspecified.

III

A LITTLE LOGIC

5

Ambiguity

The account of confusion-attribution as "semantic position-taking" suggests a strategy for answering other philosophical questions about confusion. Work out the details of a semantics, a validity criterion, that fits the bill; do the logic that is necessarily left undone by someone who attributes confusion by simply saying the subject "thinks A is B," or "does not know the difference between A and B" (another semantic use of psychological language). Then see what philosophical implications or philosophical presuppositions these semantic details have. Might work, might not. I think it is a fruitful strategy, and it is the one I will follow.

In Part Five I will argue that a plausible candidate for this "semantics of confused thought" is a four-valued epistemic semantics, a lightly customized version of Nuel Belnap's "useful four-valued logic."[1] None of Fred's "Charley" sentences has truth-values in that semantics, and it does not define validity as guaranteed truth-preservation. My defense of this choice of semantics will rest, in part, on the premise that there are good theoretical reasons to support the intuition that Fred's sentences cannot be assigned truth-values by any reasonable scheme. This chapter, and the five that follow, address that premise.

It makes sense to start by weighing the pros and cons of a semantics that assigns truth-values to Fred's sentences supervaluationally, and then defines validity as guaranteed truth-preservation. Hartry Field's thoughtful study of what he calls "indeterminacy," a phenomenon I would assimilate to confusion, winds up recommending a supervaluational semantic approach, and there is no other worked-out competitor in the literature. The leading idea of a supervaluational approach to the semantics of confusion was sketched in Chapter 1 under Option (3). As applied to Fred's confusion, the idea is that Fred uses Charley, and perhaps other confused terms, to multiply-refer to both Ant A and Ant B. When both referents would

render a sentence true (false) were they single referents, the sentence is true (false). If there is a split vote, the sentence lacks a truth-value. This formulation must be a bit more careful if our aim is to state a useful validity criterion, but since the issues I want to raise at the moment all are of a general philosophical kind, rather than matters of logical detail, this crude statement of the main idea will suffice.[2]

One way to appreciate the strengths of a supervaluational semantics of confusion is to see why it is wiser to think of "Charley" as referring multiply—if one is going to think of it as referring at all—rather than thinking of it as referring singly but ambiguously. An "ambiguity semantics" could take several forms, but the common thread is that we represent Fred as sometimes using the name "Charley" to refer to Ant A and sometimes using it to refer to Ant B. Any given token of a "Charley" sentence is true or is false, and validity is defined classically. It is obvious that people use ambiguous terms all the time without thereby confusing things with one another (money banks with river banks, for example), so a confusion-attributer cannot merely be plumping for the adoption of an ambiguity semantics for Fred's language. The claim would have to be that when we say Fred is confused, we are taking an evaluative position that includes a semantic position but goes beyond semantics into epistemics—something along the following lines:

> (Ambiguity) When we (we in the story) attribute confusion to Fred, we are taking the position that (a) his reasoning ought to be appraised for validity by means of an "ambiguity semantics"; and (b) when Fred is in fact using "Charley" to refer to Ant A (Ant B), he ought to be seen as unable to access the fact that he also can use "Charley" to refer to Ant B (Ant A).

The idea driving (Ambiguity) is that even though Fred sometimes, perhaps often, commits fallacies of equivocation when he makes "Charley" inferences, we should evaluate his *logically* infirm reasoning charitably in a larger sense. We should recognize that Fred never is aware that there is a risk of equivocation. This amounts to applying an uncharitable validity criterion, one that represents Fred as equivocating left and right, while also applying an all-in charitable epistemic standard, one that counts him as reasonable in making the invalid inferences he makes.

The jargon "unable to access the fact" in clause (b) of (Ambiguity) should be left for psychologists to explain in detail. As long as there seems to be something for the jargon to mean at an intuitive level, such defer-

ence is responsible. And there does seem to be an intuitive meaning: we all know what it is like to be talking about river banks, and at some point thinking about money banks under the same label "bank," noting apperceptively what we have done, all without breaking the flow of river bank talk. (If you had never done it before, you just tried doing it now, and it was easy.) This is all I mean by being able to access the fact that you are in command of a different meaning for a word than the meaning you are exercising at the moment. It might be that referential thought is not possible at all in the absence of this ability, but common sense does not teach us anything about that, so it is possible to imagine a person who can't do it in selected cases. Therefore (Ambiguity) is at least coherent.

But it isn't very plausible, because it isn't inferentially charitable no matter how hard it tries. Let's go a step at a time. First, if Fred is interpreted as using the name "Charley" ambiguously, how exactly might he wind up committing a fallacy of equivocation?

Suppose Fred says "Charley is a strong ant." We challenge him to give reasons. He gives us this argument:

(1) Charley once carried a big pebble.
(2) Charley once carried a big twig.
(3) Ants can do what they did once.
(4) If an ant can carry a big pebble and can carry a big twig, it is a strong ant.

Therefore,

(5) Charley is a strong ant.

If Fred uses "Charley" ambiguously, he is liable to have equivocated even in the space of this little argument. The thesis (Ambiguity) gives no details about what determines when Fred is using "Charley" to refer to Ant A and when he is using it to refer to Ant B, so we must supply some. There are alternatives.

One alternative is to think of Fred as using "Charley" to refer to the big ant before him, if one of them is before him. Presumably Fred's dispositions to utter "Charley" sentences covary to some extent with the properties of whichever big ant he is looking at. One might think this consideration of paramount importance in determining the referent of "Charley" at that time, especially if one wanted above all else to make good predictions about Fred's short-term ant-manipulating behavior, based on variations in his short-term ant-beliefs. But this would not settle which ant

Fred refers to as "Charley" when neither is before him, or when he is at a distance from the ant colony, still jabbering on about "Charley." One might have a counterfactual criterion for these "remote" uses of "Charley": Fred uses "Charley" to refer to whichever big ant would be before his eyes if he were looking at the ant colony. Or one might have an "inertial" criterion: Fred is referring to the last big ant he referred to up close.

Another alternative is to think of Fred as using "Charley" to refer to whichever big ant was the target of his investigation when he confirmed the belief a sentence expresses, or the several beliefs from which he has inferred the sentence. So if Fred thinks and says "Charley once carried a big pebble" because (psychologically because) Ant B carried a big pebble right before his eyes three days ago, then even if at present Ant A is right before his eyes, Fred is referring to Ant B when he says "Charley once carried a big pebble." One might be attracted to this rule for assigning unique referents to "Charley" if one were especially concerned to have Fred's memory-based beliefs turn out true a "normal" fraction of the time. This approach would work just as well when Fred was talking about "Charley" remotely as when he was doing it up close, but it would not work well if Fred's belief—for instance his belief that "Charley once carried a big pebble"—was derived inferentially as much from judgments he originally made while interacting with Ant A as from judgments he originally made while interacting with Ant B.

Each of these alternatives is motivated by a principle of the form: "When we interpret the reference of Fred's words, it is very important that we do so in such a way that Fred's singular beliefs have property P." Property P might be "short-term predictive power." Or property P might be "accuracy, to the extent the belief is memory-based." Both candidates for property P are plausible things for an epistemologist to care about. So if we were giving a semantics for Fred's reasoning in the service of a more general theory of knowledge, either alternative would make sense. But neither alternative tells us how to interpret every one of Fred's (token) "Charley" sentences; so on either alternative we would need to legislate in unsettled cases. And, most important for our purposes here, both alternatives will invalidate more arguments than wholehearted inferential charity should allow.

Take the little argument "(1), (2), (3), (4) so (5)," displayed above. Suppose we adopt the semantic policy that Fred uses "Charley" to refer to whichever big ant is before his eyes, if one is; and when Fred's tokening of "Charley" is remote, Fred uses "Charley" to refer to whichever big ant

would be before his eyes if he were looking. If Fred gives the argument "(1), (2), (3), (4) so (5)" at some distance from the ant colony, and the big ants switch places on the surface between his first and second premise, he commits a fallacy of equivocation.

Or suppose we adopt the semantic policy of saying Fred uses "Charley" to refer to whichever big ant was implicated in his initial verification of a sentence, or of the premises from which he infers the sentence. Obviously Fred's belief in (1) and his belief in (2) may flow from encounters with different big ants. If so, his argument "(1), (2), (3), (4) so (5)" again commits a fallacy of equivocation.

Either way, we represent Fred as leaping an inferential gap he has no logical right to leap, thereby committing a trivial fallacy of equivocation. It is irrational to go around doing that. But it is exactly this kind of irrationality charge we want to protect Fred against when we take the inferentially charitable semantic position expressed in our judgment of confusion. To say that Fred has confused the ants is, as it were, to characterize him as unable to make such a mistake of logic (and to characterize him as abnormal, maybe subnormal, in consequence), just as to say the child thinks all numbers are integers is to characterize her as unable to make certain logical mistakes, because they are mistakes only if one's domain of "numbers" is more inclusive than the class of integers. Or so I have said. Must (Ambiguity) be rejected for failing to capture this spirit of inferential charity?

This is where the epistemic twist in (Ambiguity) is supposed to save the day. According to (Ambiguity), when we attribute confusion to Fred, we are plumping for an epistemic assessment of Fred's degree of semantic awareness, to accompany our proposed validity criterion. Fred, we say, ought to be regarded as unable to access the fact that he may use "Charley" to refer to Ant B (Ant A) when he is in fact using "Charley" to refer to Ant A (Ant B). No doubt it is narrowly uncharitable to represent Fred as committing trivial fallacies, but it is not irrational of him to do so, all things considered, since if you cannot tell you are reasoning illogically, you are less at fault than if you can tell; you deserve milder criticism. Maybe you do not deserve criticism at all. Fred, we say, is not at all like a person who knows there are two big ants; who for some reason consciously uses the name "Charley" equivocally, now denoting one ant, now the other; but who carelessly fails to check carefully enough, committing a fallacy of equivocation as a result. That person would be inferentially irrational. Fred isn't, because he cannot detect his equivocations.

This is hokum. It is not all that far removed from arguing that some rea-

soner is thoroughly rational because he is too stupid to notice fallacies. In case that seems an unfair caricature, think for a moment about the kind of intellectual evaluation one makes of a person who really does commit simple fallacies of equivocation because he "can't notice them." This can happen when a person simply has a feeble grasp on what his words mean. We do not, typically, conclude that the person is perfectly rational despite having just given a trivially invalid argument. A simple made-up example:

Herman reads in two different books that
(A) Flushes are desirable (from a poker manual), and
(B) Flushes are a sign of illness (from a home medical companion).
Straightaway Herman infers that
(C) Some signs of illness are desirable.

Herman does not think (absurdly) that reddening of the skin is a type of poker hand. Herman just does not fully grasp what his words mean, but he goes ahead and uses them nevertheless, sometimes governed by a "sounds like" rule. As a result, he wrongly takes an invalid argument to be valid. This, we think, is a misstep in practicing the art of reasoning, and rather a serious misstep. The defense that poor Herman is a moron and cannot help having a flimsy grasp on what his words mean has no weight.

The fact is, we do not indict Fred on analogous grounds. Fred, we say, knows well enough what "Charley" means. Attentive though he may be to this meaning, logically sharp though he may be, he will endorse the argument "(1), (2), (3), (4) so (5)." So there must be a sense in which that really is the logically sharp thing to do; there must be a validating semantics. Or so we say, when we are of a mind to characterize Fred as confused. We must conclude that (Ambiguity) fails to capture this idea. It assimilates Fred to Herman.

6

Humoring

A person can reason in a way that ignores the distinction she perceives between two things, doing so on purpose. Probably this happens most often when an unconfused person sets out to humor an evidently confused person. Example: when someone goes on at length about the Russian revolution, it is a tad pushy to insist that a careful distinction be made between two different revolutionary events, each of which contributed to the final outcome. Certain delicate inquiries require making the distinction, but people who say "the Russian revolution" in the first place seldom pursue these inquiries. It is better not to make them mad. So one plays along, using the description "the Russian revolution" with a studied refusal to mean one event rather than the other.

Suppose Jim Bob talks of "the Russian revolution," and we choose not to make a point of it. We need a strategy for deciding when to affirm and when to deny sentences containing the expression "the Russian revolution" during our conversation with Jim Bob. Actually, our problem is even more complicated. We also need a strategy for when to use pronouns which anaphorize to some occurrence of the expression "the Russian revolution," as in the sentence "The Russian revolution brought Lenin to power, *it* did not bring Peter the Great to power." And we need a strategy for when to use expressions that refer indirectly to the Russian revolution, as in the sentence "We should tell Bubba about that grim event we were discussing yesterday." We need a strategy for how to use all of these terms in subordinate clauses, as in the sentence "I think the Russian revolution was less important than you think *it* was." And so on and so on. Humoring Jim Bob will demand that we have rules for how to play the just-one-Russian-revolution game quite generally. It would be fairly easy to humor Jim Bob, so it must be possible to have such rules, at least in the sense of "have" that means "be able to act upon." I will make the assump-

tion that if we can describe an adequate strategy for using the expression "the Russian revolution" in what traditional grammarians call "simple sentences," we will have a leg up on the problem of describing a more general strategy. And I will assume that a strategy that works very smoothly in the simple case is likely to be generalizable, given enough imagination.

Humoring involves more than just playing along. One must obey a prescribed set of "do not disturb" guidelines, but without misleading or misinforming one's interlocutor any more than absolutely necessary. Humoring does not include abandoning the norms against lying. If one undertakes to humor a person who harbors a great many bizarre beliefs, and one intends to respect most of that person's distorted world view, it could be a daunting task to decide how to go about misleading the person "no more than necessary." But we may assume that in our conversation with Jim Bob we really only want to be sure that we continue to pretend there was a unique Russian revolutionary event at the end of World War I. Under that constraint, we want to do as well by Jim Bob as we can, giving him the benefit of what we actually think.

A pretty good choice of humoring strategy—for simple sentences containing the expression "the Russian revolution"—would be this: for a given simple sentence of one-revolution talk, consider two candidate "translations" into two-revolution talk. In two-revolution talk one is obliged always to distinguish "the February revolution" and "the November revolution." One translation takes "the February revolution" for subject, the other takes "the November revolution" for subject. Then use this rule: if both translations are right in your judgment, affirm the sentence of one-revolution talk, as need be, when conversing with Jim Bob. If neither translation is right in your judgment, deny the sentence as need be when conversing with Jim Bob. If there is a split verdict—one translation is tolerable but the other is intolerable, make do without ever either affirming or denying the sentence being translated.

We might affirm "The battle of Tannenberg led to the Russian revolution," since, plausibly, the Tannenberg disaster "led to" both revolutions. And we might deny "The Russian revolution led to the ascent of Yeltsin." Surely it is a reach to see that much determinism in either the February revolution or the November revolution. If the question happens to arise whether "The Russian revolution was orchestrated by Lenin," we have no option but to change the subject. There is nothing else to do without taking the further step of clearing up Jim Bob's confusion, which would be to abandon the plan to humor him.

Now let's connect this strategy for humoring Jim Bob with a use of the predicates "true" and "false." According to the "prosentential theory of truth" (PTT), the sentences "That is true" and "That is false" are "prosentences," related to ordinary, contentful sentences (and certain quantifiers) the way anaphoric pronouns are related to the noun phrases (and certain quantifiers) which are their antecedents.[1] Consider the exchange:

John: "It is raining."
Mary: "That's true."

According to the PTT, one should not think of the word "that" contained in Mary's remark "That's true" as a pronoun denoting "the thing John said" (perhaps understood to be the statement John made, or else the proposition expressed by his words). Rather, one should think of the entirety of her remark as a "prosentence," picking up its content from John's remark, which was its anaphoric antecedent. Mary is reiterating John's remark; she is not mentioning it and commenting on it. But she avoids "plagiarism" by utilizing a linguistic device that wears on its face the need for some antecedent or other in order for it to be contentful. Analogously, had she said "That's false," she would have reiterated, but with denial. When the predicates "true" and "false" are used this way, they are not "real predicates." They are not used to ascribe the property "truth" or "falsehood" to a prospective bearer of the property, such as a sentence or statement or proposition. The subject-predicate grammar of "That is true" has no ontological significance: its sole function is to "mesh" prosentences with the grammar of ordinary English sentences.

Anaphoric pronouns can have a noun phrase as antecedent, as in "I like your car. Where did you buy it?" But an anaphoric pronoun also can anaphorize back to a quantifier, as in "Something smells bad. Is it your dog?" Prosentences behave the same way. By the lights of the PTT the sentence "Everything John believes is true" should be understood as a truncation of "For everything, if John believes that it is true then it is true," where the quantifier "binds" variables in the grammatical category of sentences and the two occurrences of "it is true" are quantificational prosentences, anaphorically linked to the initial quantifier. In order to represent this sentence without distortion in a formal language, one would need to use a class of variables for which sentences, not terms denoting sentences, may be grammatically substituted, making something like "(p)(jBp → p)" well formed even though no predicate is attached to the

second occurrence of "p." A proponent of the PTT will say that an argument like (A):

Everything John believes is true;
John believes that it is raining;
So:
It is raining

is valid because the formal representation (A*) is valid:

{1} (p)(jBp→p) [Premise 1]
{2} jBp1→p1 [from {1} by instantiation]
{3} jBp1 [Premise 2]

So:

{4} p1. [from {2} and {3} by *modus ponens*]

If, on the other hand, one thinks of the predicate "true" in "Everything John believes is true" as genuinely predicative, as property-ascribing, then argument (A) is valid because (A#) is valid (the symbols are decoded below):

{1} (x)(jBx→Tx) [Premise 1]
{2} jBx1→Tx1 [from {1} by instantiation]
{3} jBx1 [Premise 2]
{4} Tx1 [from {2} and {3} by *modus ponens*]
{5} (x)(Tx→Sx) [Premise 3]
{6} Tx1→Sx1 [from {5} by instantiation]
{7} Sx1. [from {4} and {6} by *modus ponens*]

Here the variable "x" is an individual variable, ranging over propositions. (A similar argument could be constructed with variables ranging over sentences.) Belief is taken to be a relation between a person and a proposition, so "jBx" means that John believes the proposition x. The singular term "x1" denotes the proposition "It is raining." The predicate "T" expresses the property "truth." What we want is an argument for the conclusion "It is raining," rather than an argument for the conclusion "That it is raining is true." Line {4} of (A#) says "That it is raining is true." So some principle is needed, logically, to get from that to the wanted conclusion. Make the assumption that for each proposition x, there is a sentence that serves as "the canonical articulation of x." Call that sentence "Sx."[2] What we want is a "translevel" principle like the one on line {5}. Once this principle is interpolated into the argument, the wanted conclu-

sion, line {7}, follows. Or, one might let the individual variables "x" range over sentences, and have a definition of the truth predicate which has {6}, and every other instance of {5}, as a theorem.[3] Then {5} would be unnecessary. Either way, argument (A#) and argument (A*) have different forms, (A#) being somewhat more complicated.

A good test for whether the predicates "true" and "false" are being used prosententially is whether the simpler representation (A*) or the more intricate representation (A#) is required to expose the validity of an English-language argument with the "surface" form of (A). If (A*) suffices, the true/false talk is just prosential. The words "true" and "false" do not correspond to predicates in the real grammar of the sentence, they are just fragments of prosentences.

Keeping this guideline in mind, let's introduce a purely prosential use of "true" and "false" to help us humor Jim Bob. The rules are: when it is appropriate to agree with a one-revolution sentence, it also is appropriate to say "That's true" to Jim Bob in the event he utters the sentence. When it is appropriate to disagree, it is appropriate to say "That's false." When the translation test dictates that it is impossible to agree or to disagree in good conscience, neither "That's true" not "That's false" is appropriate to say; one may, however, mumble something dodgy about "neither true nor false." Samples:

Jim Bob: "The battle of Tannenberg led to the Russian revolution."
Humorer: "That's true."
Jim Bob: "The Russian revolution led to the ascent of Yeltsin."
Humorer: "That's false."
Jim Bob: "The Russian revolution was orchestrated by Lenin."
Humorer: "Well. Hmm. Not really true, you know, but not really false either."

Probably Jim Bob would behave badly at this point.[4]

If we wish to draw conclusions about the world from Jim Bob's opinions on the Russian revolution, we can do so by means of such arguments as

Argument (JB)
Everything Jim Bob has said so far today is true.
Jim Bob said today that Lenin came to power as a result of the Russian revolution,
so:
Lenin came to power as a result of the Russian revolution.

We expressly understand (JB) prosententially. It has the form (JB*):

(p)(jb SAYSTHAT p → p)
jb SAYSTHAT p1,
so:
p1.

Now we are equipped with rules for engaging in "true"/"false" talk both in conversation with Jim Bob, and among ourselves, provided the conversations among ourselves are still in a humoring mode—exercises in saying things it would be appropriate to say to Jim Bob when humoring him, though said in his absence. In so doing we never ascribe a property—"truth" or "falsehood"—to a proposition or to anything else. Indeed we make no use at all of such properties: we do not quantify over them, we do not refer to them obliquely, we do not even presuppose that they are intelligible (nor, of course, do we presuppose that they aren't).

This prosentential style of "true"/"false" talk is the natural way to use that vocabulary as an adjunct to humoring. Even if we had not stipulated that we were going to use the vocabulary prosententially, we would start doing so as soon as we tried to find a way of "talking about truth and falsehood" that was compatible with humoring. Now as we just observed, when the words "true" and "false" are used prosententially, and hence are used as devices for expressing agreement or disgreement rather than as devices for attributing properties to someone's statements, a supervaluational rule for deciding when to say true and when to say false is natural and reasonable. So we can write down a supervaluational formula and interpret it as merely a neat way to codify our humoring use of "true" and "false" when we are dealing with Jim Bob. We run no risk, provided we do not slip up and misinterpret what we are up to as "describing the principle according to which Jim Bob's statements should be classified as true or false."

In its simplest form our rule will say something like this:

> *Rule for humoring truth-talk.* When trying to decide whether to say "That's true" or "That's false" in response to something Jim Bob says in his one-revolution language, first test the sentence in question with "the February revolution" substituted for "the Russian revolution," then test it after substituting "the November revolution." Make a judgment each time whether you are ready to affirm or deny the translation. If you are ready to affirm both translations, say "that's

> true" to Jim Bob (if you wish to use true/false lingo at all); if you are ready to deny both translations, say "that's false" to Jim Bob (if you wish to use true/false lingo at all); if the translation test yields a split decision, beat around the bush, declining to say "That's true, Jim Bob," and also declining to say "That's false, Jim Bob." If cornered, say "Frankly, Jim Bob, what you say is neither true nor false."

This rule tells us what to do when conversing with Jim Bob. It does not tell us what to do when conversing among ourselves, though still in a humoring mode—that is, still trying to talk in the way that will most effectively humor Jim Bob. I take it that, with a little beefing up with such language as "talk as you would talk when trying to humor Jim Bob" inserted at strategic points, the rule can be made quite general.

It is worth repeating that we must be careful not to misconceive what we are up to in adopting this supervaluational codification of humoring truth-talk. Grant that this is indeed the best policy to adopt when one wishes to mix truth-talk with humoring. It does not follow that the most plausible semantics for Jim Bob's language is one in which truth and falsehood are posited as semantic values of Jim Bob's sentences, to be ascribed supervaluationally, with "validity" for Jim Bob's arguments characterized as truth-preservation. In classical semantic theories a closely related cluster of properties and relations—truth, reference, truth-of (as when a predicate is "true of" an object), and a few others—are manipulated into an explanation of argument validity; often an explanation of validity as truth-preservation under all circumstances. In the theory language of this kind of semantics, the grammatical predicates "true" and "false" express properties, and simple predications of the form "S is true (false)" are used to ascribe those properties. So used, the predicates "true" and "false" are very "real," and certainly are not mere fragments of prosentences. A prosentential account of truth-talk would be quite mistaken as an account of this theoretical use. Keeping this firmly in mind, let's get to the main point of this chapter.

A confusion-attributer takes a semantic position. It is an inferentially charitable and paternalistic semantic position. But a confusion-attribution in ordinary, nontechnical language contains few semantic details; we must try to fill in those details in a way that captures the idea that the subject's inferences should not be invalidated merely because of the confusion. Is a supervaluational semantics the right choice? Let us set to one side the humoring use of "true" and "false." Let us not mistake the naturalness of

this humoring use for evidence that a supervaluational semantics would be the best choice to fill in the details of a confusion-attributer's semantic position. Are there, then, compelling considerations either for or against that semantic choice?

At first we will have to make do with intuitive considerations. In order to get at what may be the most compelling of these intuitive considerations, it will be helpful to ask a preliminary question. We (we in the story) have found a way to converse smoothly with Jim Bob (smoothly except possibly for the tense moments when we tell him that some thesis about the Russian revolution which he thinks is perfectly clear-cut is neither true nor false). We have a principled way of deciding when to affirm things expressed in one-revolution language and when to deny them. We have a use for one-revolution language, and it coincides closely with Jim Bob's. We can think one-revolution thoughts. We can engage in one-revolution reasoning. So why aren't we as confused as Jim Bob?

The first step toward an answer is to reformulate the question. We—author and reader of these words—attribute confusion to Jim Bob. We say he confuses two different revolutionary episodes. We do not attribute confusion to the humorer, to "we in the story." Why the difference? Given my account of confusion-attribution, the answer must be that we do not think Jim Bob illogical if his inferences blur the distinction between the two episodes, whereas we do think the humorer illogical if she blurs the distinction. There is nothing to be said about why we do not think Jim Bob illogical that has not already been said in connection with Fred's confusion; their cases are parallel. What needs to be explained is why we do not think inferential charity toward the humorer is in order. After all, the humorer converses in, communicates effectively in, and even makes true/false judgments in the same one-revolution language Jim Bob uses.

But the humorer does not figure out what to say in Jim Bob's one-revolution terms in anything like the way one ordinarily figures out what to say in one's language. To put it in terms of a familiar philosophers' distinction, we ordinarily reason entirely within an object language. We do not derive properties of sentences within a metatheory, conclude that a given sentence has a property in virtue of which it is assertible, and only then assert it. But that is exactly what the humorer is doing. The humorer is making decisions about what sentences to assert, or deny, or withhold assertion and denial of, by reflecting upon the "best translations" of those sentences into another language, evaluating those translations, and applying a rule which dictates what to assert or deny given this sort of information about a

sentence. The humorer's reasoning is metatheoretic, and it is conducted in a metalanguage.

When we appraise the humorer's reasoning, the patterns of inference to be appraised are very different from run-of-the-mill inferences in Jim Bob's language. Indeed, they are not in Jim Bob's language at all. The final steps of a typical inference might be:

(1) The sentence "The Russian revolution led to Lenin's ascent to power" is translated by two sentences each of which I endorse.

So, applying my rule for assertibility,

(2) I may assert the sentence "The Russian revolution led to Lenin's ascent to power."

(3) Conditions exist in my conversation with Jim Bob that make it desirable to express a view on what Jim Bob calls "the connection between the Russian revolution and Lenin's takeover." The sentence "The Russian revolution led to Lenin's ascent to power" expresses such a view.

So, by (2) and (3),

(4) I should assert "The Russian revolution led to Lenin's ascent to power."

[Output of practical inference] ⇒ "You know, Jim Bob, the Russian revolution led to Lenin's ascent to power." [This said by the humorer].

We have no good reason to appraise such patterns of inference as this by any weak-kneed, charitable criterion of argument strength. So there is no basis for a semantic position recommending as the "fair" criterion a confusion-tolerant one. Therefore it would be incorrect for us to describe the humorer as confused.[5]

On the other hand, if the humorer somehow loses sight of what she is up to and starts making inferences in Jim Bob's idiom exactly the way Jim Bob would do, we will have every right to propose a confusion-tolerant validity criterion for her, too. If we do choose that semantic position, then we will be treating her as just as confused as Jim Bob. We will be treating her as having come to "believe in the Russian revolution, that earthshaking event which overthrew the Czar and installed the Bolsheviks." *Folie à deux.* At least it would be if one called Jim Bob's world view delusional, rather than ignorant.

With that mess sorted out a bit, we can draw some helpful conclusions. When one regards a person as confused, one will not think of the person's sentences (the ones containing "confused terms") as fully intelligible. This

does not prevent one using those sentences oneself (when humoring), fully understanding what one is doing. Nor does it prevent one having a fully intelligible use for the words "true" and "false" in connection with those sentences (again when humoring). When one is humoring Jim Bob, one's rule for deciding when to affirm or deny one of Jim Bob's confused sentences, and one's rule for using the words "true" and "false" in connection with them, requires that one find the two best translations of a one-revolution sentence fully intelligible, and it requires that one find the sentences of the metalanguage, in which one reasons out what to say to Jim Bob, fully intelligible. But it does not require that one find Jim Bob's sentences fully intelligible.

And, in a sense, Jim Bob's one-revolution talk is *not* fully intelligible. If we were to utter one of Jim Bob's one-revolution sentences *in order to make a statement about how the world is,* we would be making a very poor choice, and we know it. The fact that we can find a reasonable policy for uttering that same sentence in order to humor Jim Bob does not affect this point. One-revolution talk does not correspond neatly to reality, does not fit the world tightly, even when reality cooperates to the limit of its ability. That is, even if both revolutions had the property P, we cannot use the sentence "The Russian revolution had the property P" to state a fact about the world. (Remember, a sentence like "The Russian revolution had the property P" is not equivalent to the conjunction "The February revolution had P and the November revolution had P"; it is not a vehicle for making the perfectly intelligible conjunctive point.)

It seems to follow straightaway that the confused sentences of a confused person cannot be evaluated as true of the world or as false of the world. Those sentences simply do not fit the world tightly enough to be true of the world or false of the world. And from this it seems to follow straightaway that one ought not adopt a semantics for the confused person's language that deploys Truth and Falsehood, the properties, as semantic values. A supervaluational semantics is just one of many on that blacklist. Call this the loose-fit thesis.

The loose-fit thesis is intuitively plausible. Since it bristles with metaphor, we ought to test it as concretely as possible. And any metaphor-laden intuition has limited weight until a little theory gets behind it to unpack the metaphor. We must try to find some. It will help us get on with both these items of business if we return to Fred and his ants, since his linguistic history is a entirely a matter of (imaginary) record, whereas we really don't know when, or how, Jim Bob went astray.

So, suppose we have been humoring Fred for a time, agreeing and disagreeing with him in "Charley" language, using the same supervaluational rule the humorer used when humoring Jim Bob, with the obvious changes made to suit Fred's confusion. Fred leaves the room, and we immediately put our judgmental hats back on. We could think critically and judgmentally about Fred while humoring him, provided we did not mix up the two activities. But it will be simpler to imagine that we switch from one role to the other. We happen to know that Ant A is asleep down deep in the ant colony, while Ant B is up top, also asleep. Fred is in the next room, not looking at either ant. "I bet Charley is asleep," he says. "It is rather warm in the room and that makes ants sleepy."

If we were humoring Fred, the best response to "Charley is asleep" would be "That's true." But that is not the issue. The issue is whether Fred, in uttering the sentence "Charley is asleep," is right about the way the world is. It does seem plausible to say Fred's "Charley" talk does not "fit the world tightly." Furthermore, it seems plausible to say that because Fred's "Charley" talk does not fit the world tightly, when Fred says "Charley is asleep" he is neither right not wrong about the way the world is, and one ought not ascribe truth-of-the-world to the sentence "Charley is asleep."

Now let's try to scratch out a little theory to put behind all this plausible seeming. Two lines of thought suggest themselves:

(J1) Fred believes in this big ant Charley, but Fred has confused Ant A with Ant B. He has introduced the name "Charley" to name the "single big ant" he believes in as a consequence of his confusion. But really there is no such ant as Charley.

And:

(J2) Fred believes in this big ant Charley, but Fred has confused Ant A with Ant B. He has introduced the name "Charley" to name the "single big ant" he believes in as a consequence of his confusion. But really there are two Charleys.

The letter J is for "judgmental"—meant to contrast with "humoring." Both (J1) and (J2) describe Fred as though he has a thought-myth in which there is this big ant character he has named "Charley"—in the spirit of Option (4) from Chapter 1. I pick this way of talking about Fred because it is natural, and it makes for a quick formulation of the "no Charley" versus "two Charleys" alternatives. It is a natural way to talk about

Fred because we have the linguistic practice of using the names of mythic or fictional characters, even when we do not believe the stories and are not helping tell them. This thought-myth formulation is not essential to the main point of (J1), which is that there really is no big ant Charley, and is not essential to the main point of (J2), which is that really there are too many candidate big-ant-Charleys. Recalling some terms of art from Chapter 1, (J1) emphasizes the unreality of Fred's "Charley-character," whereas (J2) emphasizes that Fred's Charley-thoughts are "due to encounters with" two different big ants. It would be consistent to combine these criticisms, saying that Fred has this "Charley the big ant" character in his thought-myth about the ant colony, and it is mythic and unreal, because his "Charley the big ant" thoughts are due to encounters with two different real ants.

It is helpful to separate (J1) and (J2) because our next step of theoretical speculation invokes them as premises in different arguments. Suppose we accept (J1). Then we might reason this way: "Given that there is no Charley, when Fred attempts to say of Charley that he has some property, Fred is neither right nor wrong, his statement is neither true nor false. That is because Charley would have to really be there in order to have, or lack, properties." Evidently this reasoning supposes that when Fred says "Charley is asleep" he is "attempting to say of Charley that he has a property." But that is plausible. Now assume we accept (J2). We might reason this way: "Given that there are two Charleys, when Fred attempts to say of Charley that he has some property, Fred is neither right nor wrong, his statement is neither true nor false. That is because if a person attempting to ascribe a property to an object fails to pick out a unique subject of predication, she is neither right nor wrong."

These are just the two prongs of Strawson's intuition about singular reference-cum-predication, applied here to a name rather than to a definite description. Our speculations are starting to take a nice, neat shape. Fred's Charley-sentences lack both truth and falsehood for familiar Strawsonian reasons. Failure to satisfy the presuppositions of singular reference—existence or uniqueness, take your pick—is exactly the sense in which Fred's Charley sentences do not "fit the world tightly enough." This gives us some theory to back up the intuition that, as a rule for ascribing truth and falsehood to Fred's sentences, supervaluation is a bad rule. And it helps us see why this result has nothing to do with whether supervaluation is a good or bad rule for making "true" and "false" judgments in the course of humoring Fred: a humoring policy for deciding when to say "that's true"

or "that's false" permits one to say these things regardless of whether Fred is successfully picking out a unique subject of predication when he uses the term "Charley"; the Strawsonian theory would be quite wrong if it were applied to these humoring uses of "true" and "false."

This little bit of theory cashes in the metaphor "Fred's language does not *fit the world tightly enough* for its sentences to be really true or really false" in terms of Strawsonian reference-failure. In Chapters 9 and 10 I will give some reasons for rejecting Strawson's idea. The reasons described in Chapter 9 seem to me compelling. The different reasons described in Chapter 10 seem to me less than compelling, but still fairly plausible. As a result, I believe we do not really have a good theoretical underpinning for the "loose-fit thesis." We need a different account of what I have been calling the "unintelligibility" one finds in Fred's "Charley" sentences when one is in a judgmental rather than a humoring frame of mind. If the metaphor "insufficiently tight fit between language (and thought) and the world" is the wrong metaphor, because in the end we cannot cash it in with satisfactory theory, what is the right metaphor?

Think again about Fred, out of sight of the ant colony, saying "Charley is asleep" when, as it happens, both big ants are asleep. One thing that seems right to say is that by using a "confused term" Fred represents the world inaccurately; indeed, represents it inaccurately enough that we cannot evaluate what he says as plain true or plain false. It follows that we ought not to evaluate Fred's inferences by means of a semantics that deploys truth and falsehood as semantic values, and, therefore, we ought not to evaluate his inferences with a supervaluational semantics. But that needs argument. Let us postpone the argument for a moment (until the latter part of Chapter 7), and focus instead on the switch of metaphors we have just made.

We have traded the metaphor of "insufficiently tight fit between a person's language and the world" for the metaphor of "the person himself representing the world with insufficient accuracy." Accuracy of representation is an epistemic concept; for example, the opinions of a person who represents things inaccurately ought to be trusted less, other things being equal, than the opinions of a person who represents things accurately. Therefore the testimony of a person who represents things inaccurately is comparatively weak grounds for a claim to know something; you don't want to defer to the views of inaccurate thinkers when you are challenged on your assertions.[6]

How might we unpack the metaphor of Fred failing to represent the

world with sufficient accuracy for us to evaluate his sentences as true or as false? To see the direction we should look to, consider the following fact about Fred's language and thought. We may suppose that the properties of Ant A and of Ant B are substantially stochastically independent (although I have built a certain amount of correlation into the story by having one ant leave whenever the other one shows up, and there would be further correlation just because both are ants). Frequently, however, Fred will affirm or deny "Charley" sentences on the basis of observation of just one of the two ants. Indeed, it is rational of him to do so, since he thinks the big ant he is observing is the only one there is. Therefore we ought to expect Fred to be blindsided with some regularity—the big ant he has failed to observe when testing a given sentence will lack the property he has concluded "Charley" has. This will happen even though Fred is following what is rational practice for a speaker of his (confused) language.

If Fred's sentence "Charley is asleep" can be evaluated as true or as false—for instance, according to a supervaluational policy—then there are real, objective likelihoods (truth-frequencies) for the occurrence of such "facts" as *Charley is asleep,* and Fred is a bad judge of these likelihoods. This line of thought quickly leads to the conclusion that if Fred's "Charley" sentences are truth-valuable, then Fred makes *judgments of the weight of evidence* that are very poor indeed, in the precise sense that they misstate the objective conditional likelihoods. This is harsh criticism of Fred's inferences. Inferential charity compels us not to be so harsh, and the only way to manage that is to refrain from evaluating the likes of "Charley is asleep" as true of the world or false of the world. The less technical, more intuitive, way to make the point is just to say that we do not think there are such facts as *Charley is asleep,* because if there are, *much of Fred's fully rational thought is objectively wrongheaded*. Our common sense about the concept "fully rational" tells us that, whatever some radical skeptic might believe, things just can't be like that.

This line of thought is fundamental to understanding why we think of a confused person as inadequately representing the world, and to understanding why we are reluctant to evaluate confused utterances as simply true or simply false. Let's work through it with more care.[7]

IV

TRUTH-VALUING

7

Calibration

Suppose Fred sets out to confirm or disconfirm a sentence of the form "Charley is F." Assume Fred "confirms" the sentence "Charley is F" by carefully "identifying Charley," and with equal care determining that the predicate F applies. When he finishes his investigation, Fred is very confident that Charley is F, as any of us would be in his shoes. Let's use the language of subjective probabilities to describe Fred's level of confidence. We'll say: "The sentence 'Charley is F' has high subjective probability for Fred."

The thought here is that just as a person's beliefs can be described by saying "So-and-so takes it to be true that p," a person's degree of confidence can be described by saying "So-and-so takes it to be likely 0.4 that p." This should not be understood to imply that the person has made an explicit judgment that the probability of the proposition that p is 0.4, just as "So-and-so takes it to be true that p" should not be taken to imply that the person has made an explicit judgment that the proposition that p is true. It would be possible to make the points we need to make without using the concept of subjective probability, but by using it the points can be made in less roundabout language.

Now assume Fred's ant-colony sentences are truth-valuable. That is, assume those sentences can be classified as true of the world or as false of the world, at least when the world is sufficiently cooperative. We need to have in mind some consistent semantic rule for assigning truth-values, so let's adopt a supervaluational rule. The sentence "Charley is F" counts as true just in case it would be true if "Charley" denoted Ant A and also would be true if "Charley" denoted Ant B. The sentence counts as false just in case it would be false if "Charley" denoted Ant A and also would be false if "Charley" denoted Ant B. If there is an Ant A/Ant B "split" with respect to F-ness, for instance if Ant A is F but Ant B isn't, the sentence "Charley

is F" is neither true nor false; the world is being uncooperative. If we are going to ascribe truth and falsehood to Fred's sentences at all—and, as will emerge in a moment, we must do so to talk about calibration—supervaluation seems the best bet. It makes essential use of the idea that the name "Charley" multiply-refers, an intuitively attractive notion (recall the discussion of Chapter 1).

Psychologists have adapted the concept of "calibration" from instrument engineering and applied it to people. A person is said to be poorly calibrated when her subjective probability for some event differs substantially from the objective probability of the event; the person is well calibrated when her subjective probability for an event is close to the objective probability. We'll need a modified version of this concept: a person is poorly calibrated for a given sentence when her subjective probability for the sentence (her degree of confidence that the sentence is true) differs substantially from the objective probability that the sentence is true; she is well calibrated if there is a good match. Thus defined, calibration is relative to a given sentence, although we can speak loosely of a person being well calibrated or poorly calibrated *simpliciter,* when it is clear what family of sentences is meant, and when the person's calibration level does not differ much from one of those sentences to another. Obviously, the concept of calibration is not defined unless the sentences in question have truth-values. Those truth-values can be assigned in a gappy fashion, by supervaluation. What matters is that there be circumstances in which a sentence is true and circumstances in which it is not, so that there can be a probability of the true-making circumstances arising versus a probability of the true-making circumstances not arising.

Poor calibration, though related to truth in the way just mentioned, is not the same thing as "being wrong about something." A person could be right about everything and still be poorly calibrated, or she could be wrong about everything and still be (rather) well calibrated. But calibration is a type of "accuracy of representation." Thermometers do not represent the world the way people do, but an analogy with broken thermometers is not too misleading: when a broken thermometer reads 74 degrees, and the temperature in fact is 74 degrees, the thermometer is in a sense just as inaccurate as it would be if the temperature were 38 degrees, because it is poorly calibrated. The thermometer is "right," of course, but that's a joke.

As long as we assign truth-values to Fred's sentences, Fred is poorly calibrated. Even if he happens to be "right," that is, even if both big ants hap-

pen to be possess the property he ascribes to "Charley," he still is poorly calibrated. That is because the goodness of Fred's calibration depends to some extent on the stochastic correlation between Ant A being F and Ant B being F.

For example, suppose F-ness is "having the sniffles." Fred tests Charley for the sniffles, using standard antkeeper's methods that need not be recounted here. Fred happens to have grabbed hold of Ant A, who does have the sniffles. So does Ant B, making Fred's belief "Charley has the sniffles" a true one by supervaluational lights—even though Fred has not checked Ant B at all. The "rightness" of Fred's belief is to some extent an accident, in exactly the way the "rightness" of the broken thermometer is an accident. This affects the goodness of Fred's calibration. The probability that Ant A and Ant B coincide as snifflers or nonsnifflers is less than perfect; indeed it may not be very high at all—though it would be fairly high if, for instance, ants seldom sniffle, or if sniffling is highly contagious among ants living near one another. We may assume that the objective probability that "Charley has the sniffles" is true will fall short of Fred's very high subjective probability for that sentence. Perhaps it falls far short.

This is a general point; nothing special about the example I chose. Ant A and Ant B are not perfectly stochastically correlated even for "universal ant properties," such as "having six legs," since certain experiments too horrible to mention can cause one ant to have fewer legs than another. Most ant properties are even less likely to be shared than these universal ant properties are. So when Fred does a good job of "checking Charley for F-ness," and feels confident in his affirmative opinion, his subjective probability usually will run rather higher than the objective probability that "Charley is F" is true.

Exceptions may be expected to occur if Fred makes a lengthy study of the question whether Charley is F, returning again and again at random intervals to examine Charley. In these circumstances it seems likely that Fred will wind up examining both Ant A and Ant B at one time or another, so the weak stochastic correlation between Ant A being F and Ant B being F will have much less impact. Assuming this sort of exceptional diligence is indeed exceptional, Fred is poorly calibrated.

Or so we must conclude, if we regard Fred's "Charley" sentences as capable of possessing truth-values. Fine. Now, why does this state of affairs give us a reason to hold that Fred's sentences *cannot* be truth-valued? Why should we not go right ahead and call some of his sentences true and others false, perhaps according to a supervaluational rule, while adding that to

do so is a bit of a joke, the way it is a joke to call the broken thermometer "right"?

Ask yourself why it is a joke to call the broken thermometer "right," and why it would be a bit of a joke to call Fred "right" when he says "Charley is F" on the basis of his observation of just one of the big ants, although it happens, by accident as it were, that both big ants are F. The answer is that ordinarily attributions of truth have certain broader epistemic implications, or (a slightly weaker claim) they lead to certain epistemic expectations. Usually one expects that a person who has many true beliefs is a person whose advice and opinions can be trusted. One has similar expectations about thermometers. We are obliged to sacrifice this linkage in Fred's case, since even if he holds many "accidentally true" beliefs "about Charley," he is not trustworthy.

This is easiest to see in connection with the expectations "about Charley" *Fred* should have in light of his own "Charley" judgments. When it happens that both Ant A and Ant B are F, so that Fred's judgment that "Charley is F" is true according to a supervaluational rule, the sentence will be less persistently true than one might expect—assuming most ant properties are transient. For example, if Ant A has just come down with the sniffles, whereas Ant B has been sniffling away for several days and is about to get well, Fred might examine Ant B (or Ant A, it doesn't matter to the point) and observe that "Charley has the sniffles." Fred may know that a sniffling ant on average keeps sniffling for 2.63 days (a fact obtainable from all standard texts), but he will be sorely disappointed; "Charley has the sniffles" will cease to be true in a jiffy, as soon as Ant B stops sniffling. On average, "Charley has the sniffles" will be less persistently true than Fred's background knowledge leads him to expect.[1]

So Fred should not trust his own best estimates to the extent we can assume he does trust them. We should not trust them either. What Fred *reasonably believes* about the world, what he believes on the grounds that he has followed "best practice" in making his judgments, pulls away from what one can reasonably believe about the world on the basis of Fred's testimony. There is a whiff of paradox here: Fred (and the rest of us) ought not to believe what Fred ought to believe. Of course the "paradox" is easily defanged. In one sense, Fred ought to believe what he learns by applying his most rational investigative practices; in another sense, Fred (and the rest of us) ought not to believe what these practices tell him concerning "facts" of the *Charley is F* variety. But the only way there can be "facts" of the *Charley is F* variety is if sentences such as "Charley is F" can be

truth-valued. If we decline to truth-value such bits of language, no "paradox" remains.

This argument is conclusive against truth-valuing those parts of Fred's (or Jim Bob's) language in which confused terms are prominent. But it is rather a sweeping argument; we would be wise to find a special case of the argument sufficient for our purposes, but with somewhat narrower, and thus sharper-edged, premises. Here is one that will do:

Remember the context of the present discussion. We are working out the details of the inferentially charitable semantic position which a confusion-attributer is tacitly proposing. For practical reasons I have narrowed the scope of that inquiry, focusing on finding a suitably charitable monotonic validity criterion. But, really, a confusion-attributer is taking a wider position. It's fun to quote oneself, so let me simply (nearly) repeat the formulation of Chapter 4:

There is more to a person being inferentially rational than that the person gives valid arguments. Many arguments are intended only as nonmonotonically strong, that is, the premises are intended to be weighty considerations in favor of the conclusion, with the understanding that this weight can be overridden by other considerations that might pop up. Therefore the scope of the inferential charity toward Fred which a confusion-attributer is proposing must be broad. The semantic position taken is that despite his confusion of Ant A with Ant B, (i) Fred can reason completely logically, and, more broadly, (ii) can make completely reasonable assessments of the weight of evidence (though perhaps he cannot make completely reasonable assessments of the truth, the credibility, or the accuracy of evidence). The language "assessment of the weight of evidence" is not meant to have a metatheoretic flavor; it is meant to be a quick way of referring to nuanced conclusory talk—such talk as:

"Blah, blah, blah; *and so, surely,* bluh,"
or
"Blah, blah, blah; *and so, likely,* bluh,"
or
"Blah, blah, blah; *and so, maybe,* bluh."

Fred, one maintains, can get this nuanced conclusory talk right, just as right as we can.

Now, if we make semantic decisions that allow Fred to be poorly calibrated—if, for instance, we assign truth-values to Fred's sentences supervaluationally—Fred will be poorly calibrated for conditional probabilities

as well as for unconditional probabilities. That is because when "Charley" sentences have truth-values, there are real likelihoods of their truth and of their untruth. Here is an example:

Suppose Fred has made a study of the correct methods for diagnosing bad head colds in ants. He "finds Charley" and applies these methods. They require him to confirm the presence of "the three cardinal symptoms of ant head colds," so Fred dutifully checks off:

(a) Charley is sniffling.
(b) Charley is grumpy.
(c) Charley will not eat a raisin.

Fred concludes that it is very likely that

(d) Charley has a cold.

Assume that the experts agree that there is only a strong nonmonotonic inference from the presence of the three cardinal symptoms to the presence of head-colds; moreover, ants differ somewhat in "typicality" in this respect. Fred knows all that, and makes his best judgment. It does not matter to this argument which big ant Fred examines, but for concreteness assume it is Ant A.

Put in the concise lingo of subjective probability, Fred's subjective probability of (d) is very high. This is due in part to his confidence in the diagnostic adequacy of the symptom-cluster he has identified in Charley. To put *that* in the lingo of subjective probabilities, Fred's conditional subjective probability of (d) given the conjunction of (a) and (b) and (c) is very high. This conditional subjective probability can be compared with the objective conditional probability of (d) given the conjunction (a) and (b) and (c).

Again for simplicity of argument, assume for a start that both Ant A and Ant B are quite typical of ants in the connection they exhibit between the cardinal head-cold symptoms and the presence of a head-cold. Write the conditional objective probability of (d) given the conjunction of (a) and (b) and (c) as P(d/a&b&c). Suppose the conditional objective probability of Ant A having a cold, given that Ant A has the symptoms, is 0.8. Make the same supposition about the conditional objective probability of Ant B having a cold, given that it has the symptoms. Assume Fred has about a 0.8 subjective conditional probability that "Charley has a cold," given that "Charley has the symptoms"—that is, given (a), (b), and (c). If Fred were using the name "Charley" to denote just Ant A, or to denote just Ant B,

his subjective conditional probability would match the objective conditional probability. He would be well calibrated for conditional probability (in one instance).

But he isn't using the name "Charley" that way. Moreover—as we may assume—Ant A and Ant B are not perfectly stochastically correlated with respect to any of the properties "sniffling," "grumpy," "spurning raisins," and "having a cold." Therefore P(d/a&b&c) must be less than 0.8, perhaps substantially less.[2] Fred's judgment of the weight of the evidence (a), (b), and (c) for the conclusion (d) is wrong (high). If we were to relax the typicality assumption, the point would stand, except that atypical ants would throw anyone off, confusion aside. But Fred would be thrown off worse.[3]

When a reasoner is poorly calibrated with respect to conditional probabilities, when she makes judgments of the weight of evidence that do not square with actual conditional likelihoods, we are accustomed to explain the fact by pointing to some defect in her methods of inquiry. She must be reasoning in ways we do not approve; after all, our methods of inquiry, our inductive practices, are good. But this is exactly what we must not do with Fred, as long as we are urging, charitably, that he is "reasoning with the best of us."

But, as long as we truth-value Fred's "Charley" sentences, Fred will turn out poorly calibrated with respect to conditional probabilities. A *modus tollens* here gives us the conclusion that, as charitably minded confusion-attributers, we ought never to truth-value Fred's "Charley" sentences. This radical surgery will save Fred from being poorly calibrated by making "poorly calibrated" undefined for him—undefined because nothing is allowed to count as a true-making circumstance for a sentence like "Charley has a cold." Then there will be no such thing as a frequency of occurrence of such true-making circumstances.

For ease of reference later on, let us call this "the calibration argument." A corollary is that we should not say that confused people "fail to represent the world accurately." The real point is that if we were to recognize "facts" of the Charley is F variety, then the people in question would stand convicted of representing the world inaccurately, although then we would no longer be characterizing them as confused. As long as we do characterize them as confused, the element of inferential charity inherent in that characterization spares our subjects the label "inaccurate thinker," but the cost is this: they speak and think in a language—Charley-talk being an example—which does not express facts. If valid reasoning can be conducted

in this language, as I will argue it can, then the relation of "really following from" that a conclusion can bear to its premises must be explainable independently of the property of "really being so," or "being a fact," that an assertion might have.

Notice that we can help ourselves to the calibration argument only by conceiving of the concept "accuracy of representation" as a statistical concept. Tradition in the philosophy of language treats language/world relations, or thought/world relations—such relations as "aboutness," "denotation," "satisfaction," or the more epistemically flavored relations "accuracy" and "fit"—as nonstatistical. This is an error, I think. This is not the place to argue the case at length; but the fruitfulness in our inquiry of a statistical understanding of "accuracy" is worth mulling over. We won't mull any further here.

Anyhow, let us agree that we ought not assign truth-values to Fred's "Charley" sentences, even when both big ants cooperate by simultaneously having or lacking some attribute. Does it follow that we cannot use a supervaluational validity criterion to flesh out the idea that Fred is inferentially rational? Does it follow that we cannot use any validity criterion formulated in terms of the Reference/Truth/Truth-of cluster of semantic concepts? Consider a straightforward supervaluational validity criterion:

(SVV) Let "interpretations" be as in standard semantics. A "determinate assignment of Referents to singular terms" assigns one or the other big ant as Referent of "Charley," and for any other singular terms likewise makes some unique assignment of Referent. Then we say: an inference is Valid if and only if for every interpretation, and every determinate assignment of Referents to singular terms on that interpretation, it is not the case that all the premises are True (relative to that interpretation and assignment) but the conclusion False (relative to that interpretation and assignment).

Classifying Fred's arguments as valid or invalid according to (SVV) does not oblige us ever to describe a "Charley" sentence as True or as False. We can embrace the result of the calibration argument, hold that "Charley has a cold" never expresses a fact, and still apply (SVV) to arguments in which "Charley has a cold" occurs as a premise, or as the conclusion. All we need do is allow that predicates of Fred's language, for example "has a cold," can be True-of Ant A or True-of Ant B. That, of course, is not controversial.

Let us take a step back, and consider what we hope that (SVV), or some other validity criterion, will do for us. In attributing confusion to Fred, we

represent that he is fully inferentially rational, although we expect that in order to characterize his inferential rationality we will need to analyze his reasoning by somewhat nonstandard logical methods. There are two separable elements of inferential rationality.

An inferentially rational thinker must give, and must endorse, arguments we recognize as logically correct, at least for the most part, without messing up by giving or endorsing logically bad arguments, again for the most part. The person's inferential practice must be up to snuff. Call this the UTS element in inferential rationality.

But an inferentially rational person must satisfy a second requirement as well. It is much harder to formulate this requirement. Roughly, the requirement is that the person must be moved by the fact that the conclusion of an argument does or does not follow from the premises. The person must care, in principle, whether an argument is valid. The wiggle-phrase "in principle" is meant to imply that the person probably will need to have the idea of "following from" explained, carefully and clearly, and it is also meant to imply that only rarely will the person actually take the time to consider what does and does not follow from what. Usually people just plow ahead. And usually nobody ever explains the rudiments of logic to them.

Both elements are essential. If a person happens to give valid arguments for the most part, but gives them because she believes her teacher will approve, or her opponents in debate will cave in, or for any combination of such logically inessential reasons, but does not herself care whether what she concludes really follows from what she assumes, that person is not inferentially rational. A similar requirement applies to many other norms of practice; there is nothing special about logical norms here. For example, to be moral a person must do what is right (mostly), and must care, in principle, whether a prospective action is right. The same fuzziness in the meaning of "care in principle" is present in these other cases as well.

It follows that a validity criterion we have proposed as a device for fleshing out the details of Fred's inferential rationality must play two roles. It must carve out a set of licensed inferences which correspond in the main to the inferences we believe are correct, so that Fred's compliance with the criterion can serve as a measure of his tendency to reason correctly. Second, it must be such that in order to be inferentially rational, Fred ought to care "in principle" whether arguments satisfy the given criterion. That is, the criterion ought to constitute a detailed spelling-out of the idea of "follows from" which Fred ought to want to honor, at least "in principle."

If a proposed criterion lacks either of these features, it is no good for our purposes. That is not to say it is no good, period. Explicating some person's inferential rationality is a very special job for a validity criterion. It is by no means the only possible job. (Compare, for instance, the job of marking inferences as safe or unsafe for extending mathematical theories; mathematicians can be a pack of irrational fools yet the marking job is untouched.) Again, there is a close analogy with other norms of practice. In order for utilitarianism to be a good theory, it must crank out what seem to us the moral things to do; but it also must provide us with a morally weighty reason for doing them, a reason of which we can say: an agent should care, in principle, whether prospective actions have this characteristic.

Now, (SVV) does a very good job of "carving out a set of licensed inferences which correspond, in the main, to the inferences we believe are correct." The inferences it licenses are essentially those of classical two-valued logic. But it does a poor job of capturing the idea of "follows from" in a way that should have authority for Fred, which he should want, in principle, to honor. Fred can reason in ways that in fact are certified by (SVV) without having any idea that there are two big ants, and in general without being able to grasp the semantic features an argument is supposed to possess in order to be valid according to (SVV). But we cannot maintain that he should care whether he has done so.

The only word Fred might use to "say something about" Ant A and Ant B is the name "Charley," or an anaphor depending on it (or some cognate, if he has any). But Fred's "Charley" sentences do not even express facts, they are not truth-valuable. So Fred has no means by which he can know, grasp, realize (and so on for other cognitive success verbs) any fact about Ant A or Ant B. That much follows directly from the calibration argument.

Now, (SVV) is like a classical semantics in that whenever it is applied, some "domain" of objects is specified. It is from this domain that extensions of predicates and referents of singular terms are draw, and it is this domain, or set-theoretic constructions from it, that we use as a "universe of discourse" for quantified, generalized, language. We propose to draw our domains from the real things, from the world. Ant A and Ant B are in the world, and most importantly so, because we intend to pick them out as referents at certain stages in the application of (SVV). But Fred cannot know, realize (etc., etc.) that these big ants are real, or are F, or are G, or are anything, because no "Charley" predication is, or ever can be, True. In

a nutshell, Fred does not share our ontology, and cannot. But we propose to draw our domains from the stuff of our ontology, from what is real. (That is one of the virtues of (SVV) for some purposes; it works with what is real.) So there is no reason why Fred should care whether his arguments meet the standard (SVV) says they should meet. He doesn't have a full picture of the ontology we intend to draw upon when we say which real objects a term may denote or a quantifier range over. And he cannot.

The only way to change this situation is to put Fred in the know about the two big ants, and let him make some alteration in his language to reflect this new awareness. Then (perhaps after a crash course in logic so he can understand what [SVV] says) we might very reasonably expect him to care whether his reasoning squares with our validity criterion. But then, fully aware that there are two big ants and fully in possession of linguistic tools for talking about them, Fred will no longer be confused, and our game will be off.

That is the rub. Usually, when we say a person should care, in principle, that her arguments satisfy some validity criterion, we can have an expansive sense of "in principle." We can mean, for example, that if we were to educate the person about the true ontology of the world (or at least what we think is the true ontology), then she ought to want to respect the validity criterion we have formulated, on the understanding that we intend the domains of interpretation to be drawn from "the real objects." But with Fred we must mean something like this: even having his erroneous conception of what objects there are, and with no ontological reeducation, Fred should care that his arguments obey a criterion of validity based on an ontology he cannot guess is the one intended. But of course he should not.

Now suppose, contrary to philosophical fact, that Fred's sentence "Charley has a cold" were True exactly when Ant A and Ant B both have colds, False exactly when both do not. Then, one might make a speech like this:

> Fred is not able to talk, or think, in a fully discriminating way about the objects that really exist, because he can only Refer Multiply to Ant A and Ant B. But this does not show that he cannot know anything at all about them. After all, the things he says concerning these objects sometimes are True. For example, if he says "Charley has a cold," that statement might be True of the world. Surely if a person has said something True of the world, the person has a quite respectable grasp on the bit of the world in question. The person can talk and think

> about that bit of the world and get it right. This is grasp enough for it to be reasonable of us to ask Fred to care that his arguments are (SVV)-valid, with validity defined against a background ontology including the big ants, despite the fact that he is no good at discriminating them. All we need do is tell him there are certain objects he cannot tell apart, but that figure in as elements of the domains we have in mind employing. We need not actually clear up his confusion. His only problem is one of discrimination, and as we have just observed, this does not prevent him being right about the objects he cannot discriminate. Of course, if it were not possible for Fred to use his Charley lingo to say, and think, True things, then the situation would be very different. We would have no basis at all for claiming he has a grasp of how things are with Ant A and Ant B, and we would have no right to expect him to respect the supervaluational validity criterion we have proposed.

I keep going back and forth on the merit of this speech, though most often I think it has merit. If it has merit, then being able to assign Truth-values to Fred's "Charley" sentences is necessary and sufficient (other things equal) for regarding (SVV) as authoritative for Fred. By showing that Fred's "Charley" sentences ought not be Truth-valued, we cut away this necessary condition.

The considerations that tell against (SVV) as the right choice for a "logic of confusion" are very general, because the calibration argument is very general. It seems likely these considerations will tell against any validity criterion formulated in terms of the Reference/Truth/Truth-of cluster of semantic concepts, although logic is full of surprises. I am ready to be surprised here, though I am not holding my breath.

In Part Five we will meet with a semantics that does not use the Reference/Truth/Truth-of cluster of semantic concepts at all. It uses a set of "epistemic" semantic concepts. We will see, in Chapter 13, that the problem confronting semantic proposals based on the classical semantic concepts—in a nutshell, that the semantics does not have authority for Fred—is eliminated by this epistemic semantic approach. It is eliminated automatically, as it were: the semantic values employed by the epistemic semantics are designed to encode "authoritative advice." But before we can get down to that business we have some important loose ends to tie up.[4]

8

Failure to Refer

In Chapter 6 we were considering another line of thought, leading to the conclusion that Fred's "Charley" sentences should not be assigned truth-values, a line of thought that seems independent of the calibration argument. That line of thought rested on the intuitive datum that the reason Fred's "Charley" sentences cannot "really" be truth-valued even in favorable cases, such as cases where supervaluing would lead to a clean assignment of truth or of falsehood, is that Fred's "Charley" language "does not fit the world tightly enough." We even sketched out a little bit of theory to unpack that metaphor: perhaps what our intuitions are latching onto is Strawsonian reference failure. Perhaps the putative referent of "Charley" does not exist, and that is why "Charley" sentences are not truth-valuable, and that is the sense in which they do not fit the world tightly enough. In support of this idea, one could cite the fact that if we were in a judgmental rather than a humoring mode, we would be tempted to say to Fred: "Look, Fred, there is no such ant as Charley." Or perhaps the name "Charley" has too many referents, and that is why "Charley" sentences are not truth-valuable, and that is the sense in which they do not fit the world tightly enough. In support of this idea, one could cite the fact that if we were in a judgmental rather than a humoring mode, we would be tempted to say to Fred: "Look, Fred, there are two Charleys." Or perhaps both things are true: there is no such ant as Charley *in the sense that* there are two Charleys.

This is an attractive line of thought. But there are good reasons to reject it. The reasons why we should reject it may be of interest quite apart from the question how to analyze confusion, since they also are reasons to reject a core element of the theory of descriptions, at least in Strawson's form. So I will lay out the details as best I can see them, in this and the following two chapters. From the point of view of our study of confusion, the bot-

tom line will be that an intuitively attractive reason for declining to truth-value confused language is, in fact, a bad reason. The calibration argument stands alone as a good reason.

I want to begin with a statement of what I will call "the theory of descriptions." In it I will want to emphasize elements of Strawson's insight that differ somewhat from elements standardly emphasized. As a result, the ensuing argument will be one of those "if the shoe fits" arguments: if what you like to call "the theory of descriptions" shares enough properties with what I prefer to call "the theory of descriptions" for the argument to go through, then you should reject what you call "the theory of descriptions." On the other hand (to mention just one example), it is possible that what you call "the theory of descriptions" traces its heritage to Russell rather than to Strawson. If so, it is possible that you will think the features of the concept of reference on which my argument depends are irrelevant to the theory of descriptions, because you think "the theory of descriptions" maintains that what appear to be denoting terms—definite descriptions, for example—in fact carry no semantic value at all and would go away in a canonical first-order-with-identity "paraphrase." Fine. The shoe doesn't fit. (Of course, that type of "theory of descriptions" would not help unpack the metaphor of Fred's sentences "failing to fit the world tightly enough.")

The Theory of Descriptions (Our Canonical Version)

It is common for people to use definite descriptions to Refer. Reference, big R, is a language-world relation. Strawson, among others, preferred to think of it as a speaker-world or writer-world relation. We blur the distinction because nothing we say will turn on making it. It may be that when people use the word "refer" in nonphilosophical circumstances, they never mean "Refer."[1] It may be that philosophers sometimes take themselves to mean "Refer" when they say "refer," and, having no good reason to do so, make philosophical errors. Nevertheless, a concept of Reference can be introduced as a posit by a philosopher engaged in semantic theory, although it does not appear possible to do this without also introducing a whole cluster of interconnected semantic concepts, all of them language-world relations. These other concepts include Truth and Falsehood, Truth-of (a relation predicates are posited to bear to sets of entities in the world, or to sets of ordered n-tuples thereof), and Being-domain-of (a semantic property a set has when quantifiers are interpreted in it and extensions drawn

from it), among others.[2] The concept of a speaker "attempting to Refer" is intelligible only as a member of this cluster of concepts.[3] I have been calling this cluster of concepts "the reference/truth/truth-of cluster." Let's capitalize "the Reference/Truth/Truth-of cluster," just to make sure we remember we are talking about language-world relations.[4]

Now here is the core of (what we stipulate to be) the theory of descriptions. Sometimes a speaker attempts to use a definite description to Refer, while using it as the subject-term of a predicative sentence, but either there is no existing thing to which the description applies, or there are too many. In either case, the speaker fails to Refer to just one entity in the world to serve as the worldly subject of a predication. As a result, what the speaker says has no truth-value. This has a two-step defense:

First, when a speaker tries to *pick out from the entities in the world* an entity of which to make a predication, this is analogous to *deciding what entity in the world* to make a predication of. If you *have not yet decided* which entity to make your predication of, but you go ahead and utter a singular subject-predicate sentence anyway, none of the entities in the world is the very entity which must possess the ascribed property in order for what you say to be true, or to be false. Your predication has no "truth-conditions." Analogously, if you *do* try to pick out some entity in the world of which to make your predication, but somehow you fail to pull it off, once again none of the entities in the world is the very entity which must possess the ascribed property in order for what you say to be true, or to be false. And once again your predication has no "truth-conditions." Any failure to pick out an entity as the very one you are making your predication of is just as destructive of your attempt at predication as is a failure resulting from simply not making up your mind.

Second, when a definite description is the expression you undertake to use to pick out an entity in the world of which to make your predication, the picking-out is accomplished by descriptive fit. There must exist precisely one entity fitting your description, or else no entity gets picked out, and the considerations of the preceding paragraph kick in.[5]

This two-step reasoning depends upon Reference being a language-world relation. To see that this is so, remember that definite descriptions have a range of uses. Sometimes a definite description is used to formulate a generalization, as in "The llama is a hairy sort of woolly fleecy goat." It is implausible that Belloc was Referring (big R) to a particular beast someplace in Peru, although someone might do that.[6] More plausibly, he was generalizing about llamas—"typically they are hairy sorts . . .," or some-

thing like that.[7] And definite descriptions frequently anaphorize to quantifiers or indefinite noun phrases, as happens in (1) and (2):[8]

(1) Any teacher who mumbles will do a poor job. The teacher should be fired.
(2) If a teacher mumbles, he will do a poor job. The teacher should be fired.

If an utterer of (1) or (2) is cautious, she will concede that there might be no mumbling teachers, and that perhaps no teachers deserve firing. In fact, definite descriptions buried inside the most godawful constructions can anaphorize to universal quantifiers buried inside other godawful constructions, as in (3):

(3) Susan is prone to argue that any teacher who mumbles will do a poor job. I am inclined to think the teacher should be fired.

These "Geach sentences" pose nightmarish problems for intentional logicians, but there is no doubt about their grammaticality. The definite descriptions in (1) and (2) look like natural-language analogues of variables bound by a universal quantifier. The definite description in (3) seems to be working that way too, at least in the surface grammar, except that a quantifier inside a wildly nonextensional context is "binding a variable" outside that context and inside another wildly nonextensional context. Whatever semantic treatment (3) deserves, it would be a mistake to think that a person who utters (3) is presupposing that some object fits the description "I am inclined to think [. . .] should be fired."[9]

The word "refer" (small r), or as we might say in old lingo "refer in its vulgar sense," is correctly used in connection with every one of these examples, even though they are a bad fit to the theory of descriptions. As just one example, consider the conversation:

Mary: "Susan is prone to argue that any teacher who mumbles will do a poor job. I am inclined to think the teacher should be fired."
Tom (having caught just the end of what Mary said): "What teacher are you talking about"?
Mary: "I was referring to those teachers Susan says will do a poor job, the mumblers."

Since Mary can easily assume that no objects fit any of the descriptions in play here, without the slightest incoherence, the theory of descriptions better not be written down with "refer" substituting for "Refer." The point of the (Strawsonian) theory of descriptions has to be that speakers

do sometimes use definite descriptions in such a way that the descriptions bear the language-world relation Reference to something, and other times speakers try to do that but fail—thereby inducing a "gap" in the usual assignment of truth-values.

Now let's get back to filling in our (candidate) explanation why Fred's "Charley" sentences lack truth-value. As I have formulated it, the theory of descriptions has two separable components. First, it holds that when a speaker attempts to supply a subject of predication by using a term to Refer to some unique item in the world, but the speaker fails to Refer to a unique item, what the speaker says is not truth-valuable. Second, it holds that definite descriptions Refer to the item they fit. The first but not the second component applies to proper names, demonstratives, and other potentially Referring devices, as well as to definite descriptions. But as I have formulated the theory, it contains no account of the conditions in which these various nondescriptive terms Refer to one item rather than another; it only provides that information for definite descriptions. Therefore, in order to apply the first component of the theory of descriptions to Fred's use of the name "Charley," we need an argument that Fred is not successfully using that name to Refer to a unique item in the world. In the spirit of Saul Kripke's account of how a run-of-the-mill proper name should be thought of as connected with a definite description used to fix the Reference of the name, let us stipulate that when Fred introduced the name "Charley," he did so by uttering the sentence: "Let 'Charley' denote the big ant." Whatever his speech-act was, it must have had essentially that effect. We will say that Fred was attempting to fix the Reference of the name "Charley" as whatever unique entity in the world is Referred to by the description "the big ant."[10] Unfortunately for Fred, his attempt at Reference-fixing misfired, because no unique entity existed fitting the description he chose.

The sentence Fred uttered in order to fix the Reference of "Charley" was not a declarative. We have two options. The ambitious option is to widen the scope of the theory of descriptions so that it applies to imperatives and other nondeclarative sentences. The unambitious option is to identify a declarative sentence which must be true in order for Fred's putatively reference-fixing speech-act to have succeeded. Let us take the unambitious option. Here is a thesis (based on another, related, idea of Kripke's):

> (Thesis K) If a person successfully introduces the name **n** into the language as a Referring term by saying "Let **n** denote the F," then in the

> vicinity of this successful Reference-fixing, the person can know the truth of the sentence "**n** is the F" merely by reflecting upon her act of Reference-fixing. (Since the person's attempt at Reference-fixing may have misfired, as does Fred's, she cannot *know that she knows* the truth of "**n** is the F" without additional information).[11]

If we wanted to give an argument for Thesis K, we would have to be ambitious after all and expand the reach of the theory of descriptions to nondeclarative sentences. But since the thesis has considerable intuitive plausibility, we will just use it without defense.

The sentence "Charley is the big ant" is a declarative, and it employs a definite description to pick out a Referent. So the theory of descriptions in our canonical form applies. Now recall these "judgmental" (humorless) comments from Chapter 6:

(J1) Fred believes in this big ant Charley, but Fred has confused Ant A with Ant B. He has introduced the name "Charley" to name the "single big ant" he believes in as a consequence of his confusion. But really there is no such ant as Charley.

(J2) Fred believes in this big ant Charley, but Fred has confused Ant A with Ant B. He has introduced the name "Charley" to name the "single big ant" he believes in as a consequence of his confusion. But really there are two Charleys.

If "the big ant" does not Refer to any big ant (the spin we get from (J1)), then according to the theory of descriptions the sentence "Charley is the big ant" lacks a truth-value. And if "the big ant" Refers to an overabundance of big ants (the spin we get from (J2)), then once again according to the theory of descriptions the sentence "Charley is the big ant" lacks a truth-value. But if the sentence "Charley is the big ant" lacks a truth-value, Fred does not know that the sentence is true, including "in the vicinity of" his attempt at Reference-fixing. By Thesis K it follows that Fred did not successfully introduce "Charley" into the language as a Referring term. Therefore, as Fred uses the name "Charley," it does not Refer at all.[12]

So the first component of the theory of descriptions has purchase: when Fred makes a predicative "Charley" statement, he is attempting to supply a subject of predication by using a term to Refer to some unique item in the world, but in fact he is failing to Refer to any such unique item. What he says is not truth-valuable.

The upshot is that our explanation why Fred's "Charley" sentences are not truth-valuable relies on the theory of descriptions supplemented by Thesis K; that is, supplemented by a principle that lays down an epistemic constraint on Reference. In the next chapter I will argue that we should reject this explanation, because we should reject the theory of descriptions. We will have to rest content with the calibration argument as our sole reason for declining to truth-value the things Fred says and thinks. But it is worth pausing to notice that we were able to complete the semantic argument of this brief chapter only by interpolating an epistemological premise. No surprise, really, since (as I pointed out in Chapter 1) the boundary between semantics and epistemology is like the boundary between Kuwait and Iraq; some imperialist in a tent drew it in by hand to make the map neat.

9

How You Convince People—Including Yourself—of the Theory of Descriptions

Philosophers sometimes defend their theories by describing a philosophical thought experiment. One is supposed to perform the thought experiment, and in the course of performing it, come to see that the theory is correct. In analytic philosophy the thought experiment sometimes consists of attempting, experimentally, to accept some combination of propositions while rejecting others, finding that this is an incoherent mental act, and then adopting a hypothesis that explains this incoherence; perhaps on the grounds that anyone who accepts some of the propositions also tacitly and (maybe) unknowingly accepts the others; perhaps on some other grounds. Thought experiments of this incoherence-testing sort are fraught with risk. The most common problem is that by the very nature of thought experimentation each of the mental acts performed in "running the trial" is performed in the cognitive neighborhood of all the others. If it happens that some of the acts also can be performed when none of the others is performed, important features of the mental act may change.

For instance, if a thought-experimenting epistemologist tries to decide whether one can coherently ascribe knowledge of some proposition to a subject who cannot answer certain kinds of questions having to do with the "subject-area" of the proposition, she may "discover" that it is incompatible with knowledge-ascription to do so. But since the method of her thought experiment obliges her to imagine raising, or entertaining, questions about the subject-area of the proposition, she may have learned only that knowledge-ascription obeys the envisioned conceptual constraint when one has explicitly raised the issue of the subject's broader expertise. When nobody has done so, it might be perfectly acceptable to ascribe knowledge of the proposition to a subject who as a matter of fact lacks the supposedly conceptually required expertise.

The theory of descriptions is an artifact of this kind of mishandled thought experiment. The two-step argument for the theory we briefly surveyed in Chapter 8 really is a late stage of a thought experiment, the stage at which the experimenter settles upon the theory of descriptions as providing the best explanation of the incoherence detectable in an incoherent combination of mental acts already performed. We philosophers are so familiar with the thought experiment that we do not need to be told to perform the early stages. We know how they come out. The question we think needs asking is why that is how they come out. In fact what needs asking is how the thought experiment starts. Here is a sketch of the thought experiment, start to finish:

One imagines using a definite description as the grammatical subject of a predication. In the course of the thought experiment, one discovers that as one is using the description, it would be incoherent to assume that no object exists fitting the description, and it would be incoherent to assume that several objects exist, all fitting the description. One then posits, as an explanation of this double incoherence, that one is "endeavoring to use the definite description to Refer to" some existing object, and to no other existing object. One's reasoning is that if one did this, and if moreover one tried to ascribe a property to one's referent, the resulting predication would be an unintelligible linguistic (or mental) act—unintelligible exactly as it would be unintelligible to attempt to ascribe a property to an object when one had not yet decided which object to talk about. Finally, one characterizes the double incoherence in terms of *awareness of a truth-value gap:* "If the assumptions I have been making actually were true, a necessary condition for my statement to have a truth-value would be unsatisfied. This [so the gloss continues] accounts for the perceived incoherence: it is just *perceived truth-value-less-ness.* Indeed [the gloss concludes] this perceived truth-value-less-ness is so obvious I cannot even intelligibly make the assumptions."

It might seem that the logic of the situation obliges the proponent of the theory of descriptions to proceed by first identifying the "Referential uses" of definite descriptions, and then performing this thought experiment on some representative samples of these Referential uses. That would require that there be an independent criterion of Referentiality, as opposed to mere small-r referentiality. But the impression one gets from he very considerable literature is that Referentiality just "presents itself." The fairest way to look at the situation is this: it isn't that Referen-

tial (big R) uses of definite descriptions sit there glowing, singling themselves out as Referential uses. What happens is that the same thought experiment in which one notices an incoherence to be explained by the theory of descriptions marks that use of a definite description as a Referring use. The characterization of the definite description as Referring is just one more explanatory posit laid down in the course of the thought experiment.

I do not believe the thought experiment confirms the theory of descriptions. Nor do I believe that the diagnosis of the "perceived incoherencies" one confronts in the course of the thought experiment should make use of the language-world relation of Reference. Reference, big R, is beside the point. In fact, the "perceived incoherencies" are perceived ungrammaticalities, in a sense I will try to make clear. I am not attaching any sophisticated technical meaning to the terms "grammatical" and "ungrammatical"; a bright eighth-grader would have every one of the perceptions of grammaticality or ungrammaticality we will be concerned with (and indeed would be inclined to use the words "grammatical" and "ungrammatical" to describe what she perceived, though this doesn't matter). It does matter that these "grammaticality" judgments are easily made without special linguistic or philosophical knowledge, since, as we know from experience, people untutored in such disciplines can be got to have the "perceptions" invoked in the thought experiment, and can be got to have them vividly. We need some preliminary examples.

When a speech-act consists of two or more consecutive sentences, it may be necessary to group them in pairs, or in longer sequences, to decide whether the sentences are grammatical, and whether in a derivative sense the speech-acts are grammatical. For example, both component sentences of speech-act (A) are grammatical:

(A) Several dogs are playing in the yard. They are having a good time.

But something is wrong with the grammar of speech-act (B):

(B) Several dogs are playing in the yard. It is having a good time.

And something is wrong with the grammar of speech-act (C):

(C) A dog is playing in the yard. They are having a good time.

One usually thinks of the second sentence in a sequence like (B) or (C) as the culprit. In (B) the pronoun beginning the second sentence should

be plural, in (C) it should be singular. Switch the first sentences of (B) and (C), and everything will be fine. Either (B) or (C) could be a sub-act of an even more complex speech-act. If we knew about the rest of that larger speech-act, we might decide there was no grammatical problem after all, though we might also think the speaker graceless. For instance, (C) might have been uttered as a fragment of:

(D) A dog is playing in the driveway. A dog is playing in the yard. They are having a good time.

Since it is easy to infer that the dog playing in the driveway is not the dog playing in the yard, the plural pronoun is grammatical enough. It remains true that when we are presented with (C) out of its larger context, we rightly judge it ungrammatical. Therefore that judgment of ungrammaticality must be "prima facie"—a judgment that on its face (C) is ungrammatical. Such a prima facie judgment does not imply that the speech-act is ungrammatical even if one takes account of the entire linguistic history in which it is embedded. In a similar way the speech-act below strikes us as grammatical:

(E) A dog is playing in the yard. It is having a good time.

But we would be uncertain how to parse:

(F) A dog is playing in the driveway. A dog is playing in the yard. It is having a good time.

Since people talking casually are liable to string sentences together as the sentences come to mind, rather than according to some such formal principle as "Nearest quantifier governs," we really do not know whether or not the meaning of (F) is "A dog is playing in the yard. It is having a good time. By the way, a dog is playing in the driveway, too." So our judgment that (E) is grammatical is prima facie, just as our judgment that (C) is ungrammatical is prima facie.

In (A), (B), and (C) we have anaphoric pronouns in the second sentence, with the (apparent) antecedents of those pronouns in the first sentence. Things foul up when there is a breakdown in this anaphoric connection—a mismatch of number, for instance. Instead of anaphoric pronouns we might have anaphoric definite descriptions, as in:

(G) There is a nickel in my pocket. The nickel is old.

No problem here; the speech-act is prima facie grammatical. The definite description "The nickel" has a suitable antecedent in the first sentence of the pair, the quantifier-complex "there is a nickel." That distinguishes the grammatical relation between the first and second sentences of (G) from the grammatical relation in (H):

(H) There is a nickel in my pocket. Some nickels are old, some are not.

Put any other grammatical sentence in place of "There is a nickel in my pocket" in (H), and the resulting sequence of two sentences will be grammatical. You could even pick a sentence to put in place of "There is a nickel in my pocket" that would contradict the second sentence; for instance you could write:

(I) There are no nickels anywhere. Some nickels are old, some are not.

This is inconsistent, but it is grammatical. The second sentence of (H) is grammatically free-standing, so it does not matter what the first sentence of (H) happens to be.

By contrast, suppose you write:

(J) There is a nickel in my pocket, and there is a nickel in my desk. The nickel is old.

This is ungrammatical, prima facie, and therefore neither consistent nor inconsistent. (J) is prima facie ungrammatical because there are too many candidates for the job of anaphoric antecedent of "the nickel." In G, the second sentence ("The nickel is old") is not grammatically free-standing, since the description "the nickel" needs an anaphoric antecedent. It gets one, and exactly one, so (G) is grammatical. In (J), the second sentence ("The nickel is old," again) is not grammatically free-standing, because once again it needs a unique anaphoric antecedent but doesn't get one.

George Wilson has pointed out that when definite descriptions are used the way "the nickel" is used in these examples, there is a close analogy with the syntax of the "E-parameters" employed by Fitch-style natural deduction systems to instantiate E-sentences.[1] When you are reasoning in a Fitch system you can instantiate E-sentences, provided you obey certain regulations. For example, suppose you have reached a point in some proof where you have the right to write down "(Ex)(Fx)." You put this sentence on a "shelf," making it clear what is "under that shelf" and what is "out from under that shelf but under another one." For simplicity, let us just use a broken line, like this:

(Ex)(Fx)
———

Provided you are obeying the rules governing choice of E-parameter, it is legal to write under the shelf an "instantiation" of the E-sentence, in which the quantifier is dropped and the variable replaced by an E-parameter, thus:

(Ex)(Fx)
———
Fα

Then suppose you have a right to use as a premise the A-sentence "(x)(Gx)." You can go ahead and instantiate:

(Ex)(Fx)
———
Fα
———
Gα

Finally, if you choose to do so, you may existentially generalize:

(Ex)(Fx)
———
Fα
———
Gα
———
(Ex)(Gx)

Looking just at the second and third inferences, you seem to be treating the E-parameter α as though it were a singular denoting term. But of course it cannot be, as the first inference makes clear. The bare information that *something* is F does not entitle you to state of some particular object, that it, that very object, is F. Analogously, the bare information that there is a dog in the yard does not entitle you to state, of some object, that it, that very object, is a dog in the yard. So if you take as a premise the sentence "There is a dog in the yard" and correctly infer "The dog in the yard is a dog in the yard," you cannot be using "the dog in the yard" as a singular denoting term. Wilson rightly identifies E-parameters as a new part of speech—new given the usual philosopher's taxonomy.

One of the rules constraining the use of E-parameters is a prohibition against reusing an E-parameter to instantiate two existential quantifiers. For example, you may not write down:

(Ex)(Fx)
———
Fα
———
(Ey)(Gy)
———
Gα
———
Fα&Gα
———
(Ex)(Fx&Gx)

—thereby foolishly inferring (say) that something is round and square from the information that something is round and something is square. You are permitted to reuse the parameter α to instantiate an existential quantifier out from under the shelf upon which "(Ex)(Fx)" rests, but not under that shelf.

This is typical of the motivations behind the various constraints on the use of E-parameters (and A-parameters) in a Fitch system. The system is supposed to provide us with sound rules for deductive reasoning, and (hopefully) even a deductively complete set of rules. The concepts of soundness and completeness are semantic concepts. So when we explain our decision to place certain constraints on the use of E-parameters by pointing out that otherwise certain invalid arguments would be allowed, we must have in mind at least a general plan for semantic interpretation of the system's language. But this semantics need not make use of the idea that predicates are "true of" classes of objects, or the idea that singular terms "denote" objects, or the idea that quantifier variables "range over" objects, or the idea that sentences are "true" or "false." The semantic metatheory of Fitch systems often is set forth in terms of just these concepts, but that choice is not forced merely by an appeal to the ideas of "valid inference" or "valid formula." To the extent that the syntax of E-parameters in a Fitch system is semantically motivated, the motivation is neutral as between Reference/Truth/Truth-of semantics, for which the concepts of denotation and truth (relative and absolute) are central, and other types of semantics. The upshot is that you can agree to be bound by

the syntax of E-parameters without thereby tacitly committing yourself to understanding the language in terms of Reference and Truth.

Fitch systems are intended to mimic fairly well the patterns of reasoning expressible in a natural language, so they need to employ grammatical constructions broadly familiar from knowledge of, say, English. The use of E-parameters is one of these English-like syntactic features. As a result, we English speakers can imagine using the language of a Fitch system, with its predicates, terms, and logical symbols suitably interpreted, to converse in a free-wheeling way, sometimes constructing arguments, but sometimes just saying whatever is on our minds.

The idea would be to keep the introduction and elimination rules for parameters, but let a parameter, once fairly introduced, serve as a subject-expression in sentences one has not bothered to prove. We would think of an E-parameter as needing an E-sentence to "support" its use, since the parameter must have got into play by E-sentence instantiation. It would be impermissible to have two or more rivals for the role of "supporting E-sentence," and it would be impermissible to haul the parameter out from under its home shelf to use elsewhere—unless it is reused to instantiate—since this would amount to treating it as a singular term with a free-standing meaning. Looked at this way, the syntax of parameters is a lifelike, though still a toy, version of the grammar of anaphoric singular descriptions in English—"that dog," "the latter cat," "the aforementioned King of France," and all the rest of those expressions to which the theory of descriptions is presumed to apply.[2]

But the theory of descriptions makes essential use of the concepts of Reference and Truth, whereas one can avail oneself of the syntax of E-parameters without interpreting one's language (the language of some Fitch system) by means of the concepts of Reference and Truth. Therefore, if the linguistic phenomena supposed to be explained by the theory of descriptions are as well explained, or better explained, as arising from the "E-parametric grammar" of certain English singular descriptions, we ought to reject the theory of descriptions and keep to the less metaphysically committed account of the data.

Now let's return to the thought experiment one performs in order to convince oneself of the theory of descriptions. This thought experiment requires one to follow certain instructions, possibly self-imposed. "Having assumed that there is no nickel in your pocket," one instruction begins, "affirm (silently if you wish) that the nickel in your pocket is old." We can represent these instructions either in the form of imperatives, thus:

(I) Make the assumption: "There is no nickel in my pocket."
(II) Affirm: "The nickel in my pocket is old."

Or we can represent the instructions by writing out deontic declaratives to which the imperatives give rise, thus:

(I*) You must assume that there is no nickel in your pocket.
(II*) You must affirm that the nickel in your pocket is old.

Let's work with (I*) and (II*), just because I have trouble with the logic of imperatives. From (I*) it surely follows that:

(III*) You must not assume that there is a nickel in your pocket;

and, moreover, it follows so obviously that it would take a very dull undergraduate not to appreciate the fact (recall that this is essential; we know it is easy to convince people of the theory of descriptions, and that must be explained). (II*) and (III*) together are an incoherent set of obligations. Why? The theory of descriptions explains the incoherence of (I*) and (II*), but not the incoherence of (II*) and (III*). If the incoherence of the instructions you are attempting to follow arises from the fact that you are obliged to refer to a nickel in your pocket while at the same time assuming that no such object exists, it is precisely the conjunction of (I*) with (II*) that will give you pause. But why would it matter to you at all if you did *not* assume that there *is* a nickel in your pocket? This does not imply that you must assume that there is no such nickel. Compare with: "You must not assume that Bill Clinton is asleep at this moment." This does not imply that you must assume he is awake.

What would explain the incoherence of the set of obligations (II*) and (III*) is this: the expression "the nickel in my pocket" in the sentence "The nickel in my pocket is old" is like an E-parameter in a Fitch system. It must be put into play by "instantiation of" the E-sentence "There is a nickel in my pocket." Or, to put the point without the analogy with E-parameters, you must assume that there is such a nickel in order to use the description grammatically, in order to supply an antecedent for an anaphor. The incoherence of (II*) and (III*) is like the incoherence of

(III*) You must not assume that there is a nickel in your pocket,
with
(III#) You must affirm that it is old.

An obligation to affirm "It is old" in reference to a nickel in your pocket—here using "reference" in its ordinary and popular sense—carries

with it an obligation to assume that there is a nickel in your pocket (perhaps by assuming some other E-sentence which plainly entails that there is a nickel in your pocket—let us not try to dig through those details). Similarly, an obligation to affirm "The nickel in my pocket is old" in reference to a nickel in your pocket carries with it an obligation to assume that there is a nickel in your pocket. The link is syntactic; the incoherence one risks is prima facie ungrammaticality.

Since (I*) implies (III*), and obviously does, although (III*) does not imply (I*) at all, this same explanation holds good for the incoherence of the set of obligations (I*) and (II*). Therefore a syntactic explanation explains *all* the "data" one can uncover in the thought experiment, whereas an explanation in terms of one's failure to refer to an object in the world only explains half.

A similar point can be made concerning the thought experiment one performs to convince oneself that "a Referring use of a definite description presupposes the uniqueness of a referent." Here the instructions you follow correspond to the deontic declaratives:

(I**) You must assume that there is a nickel in your pocket, and another one as well.
(II*) You must affirm that the nickel in your pocket is old.

Clearly, (I**) and (II*) are an incoherent set of instructions. A syntactic explanation of the incoherence goes along these lines:

Remember why Fitch systems prohibit using an E-parameter to instantiate an E-sentence, and then using it to instantiate another E-sentence "under the same shelf." One wants to avoid such inferences as:

$$(\exists x)\{Fx \,\&\, (\exists y)[Fy \,\&\, -(y=x)]\} \rightarrow$$
$$F\alpha \,\&\, (\exists y)[Fy \,\&\, -(y=\alpha)] \rightarrow (\exists y)[Fy \,\&\, -(y=\alpha)] \rightarrow$$
$$F\alpha \,\&\, -(\alpha=\alpha) \rightarrow -(\alpha=\alpha)$$

English grammar protects against these inferences in the same way, by requiring that one not leave it open what quantifier an anaphoric definite description anaphorizes to. To replicate in English our little fallacious "proof" that α is not α, our phrasings must be stilted, for example:

(a) There is a nickel in my pocket, and another one as well; so
(b) There is a nickel in my pocket, and there is a nickel in my pocket which is a different nickel;

so
(c) There is a nickel in my pocket;
so
(d) *The nickel* is a nickel in my pocket;
but
(e) There is a nickel in my pocket which is a different nickel;
so
(f) *The nickel* is a nickel in my pocket which is a different nickel;
so
(d) and (f).

I shortened "the nickel in my pocket" to "the nickel" to make all of that more nearly pronounceable, but the point is unaffected. A person can grasp the English grammatical requirement that there be no ambiguity about the anaphoric antecedent of a definite description, without ever thinking through the manifold inferential disasters precluded by that requirement. But I suspect that at least a vague sense that "one does not know how to go on" if the antecedent has been left unsettled gives the grammatical requirement more than a schoolbook bite. However that may be, the requirement is real, and violations of it result in partial unintelligibility.

Fulfilling obligation (I**) means setting up, as an assumption, line (a) of our English-language version of the bad proof, and (a) is equivalent to (b), in which the two E-quantifiers are explicit. But (II*) requires you to make an affirmation using a definite description which has not been decorated with any indicators as to which quantifier is the antecedent. You might as well say: "There is a nickel in my pocket, and another one as well. *It* is old." The incoherence of (I**) and (II*) is, once again, the incoherence of the prima facie ungrammatical.

Now consider whether we can explain the incoherence of the set of obligations (I**) and (II*) by positing a failure of Reference to an object in the world. The problem confronting such an explanation is that whether these obligations are incoherent is a prima facie matter, and changes according to what larger context for (I**) and (II*) one takes into account. It will be much easier to bring this out if we change the thought experiment somewhat, introducing an explicit diachronic element. So suppose you try to convince yourself of the theory of descriptions by imagining that you make a temporal sequence of discoveries and inferences, but setting things up so that at a certain juncture you learn that the definite de-

scription "the nickel in my pocket" does not have a unique referent. Let your imaginary learning experience go like this:

[Background knowledge]:
I know that any nickel in my pocket is bound to be an old one.
[Time T1; I learn from a reliable source]:
There is a nickel in my pocket.
[Time T2; I infer]:
The nickel in my pocket is old.

So far, so good. Now imagine your knowledge increases.

[Time T3; I learn from the source]:
There is another nickel in my pocket as well.
[Time T4; I infer]:
The nickel in my pocket is old.

At time T5, you step back and consider what you have done. What you decide is that the sequence of discoveries and inferences you have imagined making is incoherent. That seems right. Perhaps the explanation is that at time T3 you obtained information that makes it clear you must not try to use "the nickel in my pocket" to Refer, since there is no unique Referent. But you went ahead trying to use it to Refer anyway. So you uttered "The nickel in my pocket is old" on two occasions, and *now you can see that neither token had a truth-value*. It is this recognition that you are reporting with the words "What I have said is incoherent."

Now pretend you are undergoing this alternative temporal development of your knowledge:

[Background knowledge]:
I know that any nickel in my pocket is bound to be an old one.
[Time T1; I learn from a reliable source]
There is a nickel in my pocket.
[Time T2; I infer]
The nickel in my pocket is old.

So far, so good. Now imagine your knowledge increases:

[Time T3; I learn from the source]
There is another nickel in my pocket as well.
[Time T4; I infer]
The other nickel in my pocket is old, too.

This time you judge at T5 that you never messed up. There is nothing incoherent in this sequence of discoveries and inferences. But according to the previous hypothesis, there ought to be something incoherent. You have switched definite descriptions at time T4, but the discovery you made at time T3 was just the same as in the previous development. So if last time you were able to see that both your tokens of "the nickel in my pocket is old" lacked truth-value, and you reported this recognition by saying "What I have said is incoherent," why are you not now reporting that what you have said is incoherent? You still tokened "The nickel in my pocket is old" *once*. Why don't you judge at T5 that your token of a sentence at T2 was without truth-value, and judge further that what you have said is incoherent?

If we attribute the incoherence of the first temporal sequence of statements to reference failure, that explanation ought to carry over to the second sequence. But it does not. Your interpolation of a disambiguating noun phrase *after* the juncture at which you learn there is no unique referent for "the nickel in my pocket" makes all the difference. That is just what one would predict from the hypothesis that the perceived incoherence of the first sequence is due to ambiguous anaphora. What expression anaphorizes to what other expression *can be altered by what you say later;* that is one of the reasons why judgments of the grammaticality of a speech-act are only prima facie.

Here is a bad argument that once occurred to me, because I was focused on somehow using the leading ideas of the theory of descriptions to diagnose our intuitive disinclination to truth-value Fred's "Charley" sentences:

Fred uses such definite descriptions as "the big ant" ungrammatically. This is because in his language these descriptions are put into play, as it were, by instantiation of certain E-sentences. For example, if Fred sees Ant A with a twig at two o'clock, and sees Ant B with a twig at three o'clock, he will accept the generalization "There is a big ant which had a twig at two and had a twig at three." He may then say "The big ant has admirable endurance," using the definite description "the big ant" to anaphorize to the quantifier "there is a big ant." But the generalization is false. No big ant had a stick at two and a stick at three. Therefore Fred uses the definite description ungrammatically. If Fred *then* attempts to introduce the name "Charley" by the formula "Charley is the big ant," he relies upon an ungrammatical formula (ungrammatical for containing an ungrammatical description). It is this stew of (prima facie) ungrammaticality we are intu-

itively grasping when we think Fred's "Charley" sentences ought not be truth-valued.

This is a howler. One can "put into play" an anaphoric definite description with only a false sentence to supply an antecedent (assuming we are willing to apply the concepts of truth and falsehood), just as in a Fitch system one can instantiate a false E-sentence with an E-parameter, for instance when giving a *reductio* proof. The truth or falsehood of the antecedent has nothing to do with the grammaticality of an anaphor. In the thought experiment one performs in order to "confirm" the theory of descriptions, the incoherence of assuming that there is no F and still affirming that the F is G is due to ungrammaticality. But remember, this ungrammaticality results from committing oneself *not to assume a sentence containing a suitable antecedent*. In a *reductio,* for example, one *does* assume a suitable E-sentence, for *reductio* purposes. And Fred does assume a suitable English E-sentence to provide him with an antecedent for his use of the anaphoric description "the big ant." He assumes, though falsely, that there is a big ant which had a twig at two and had a twig at three (more precisely: he assumes this falsely if we are prepared to truth-value his sentences).

Let's be clear about what the arguments of this chapter show. They show that the thought experiment one performs in order to convince oneself of the theory of descriptions can be given another interpretation. On that interpretation the incoherence one encounters in an attempt to say that the F is G, while denying that there is an F, is ungrammaticality. Similarly, the incoherence one encounters in an attempt to say that the F is G, while making the assumption that there is more than one F, is ungrammaticality. This explanation of the incoherence makes no use of the idea that Reference failure is at the root of the incoherence, depriving the speaker or thinker of a wordly item about which to talk. Moreover, we should expect the incoherence to be only prima facie, since our intuitive judgments of the ungrammaticality of speech-acts are prima facie. Our "diachronic" thought experiment bears this out.

And, since Referential (as opposed to small-r referential) uses of definite descriptions are supposed to be identified as such by recognizing that Reference needs to be posited to explain the incoherence, our arguments also show that this test for whether a description is used Referentially is a poor test.

But our arguments do not show that speakers never use definite descriptions Referentially. No argument could show that. The Reference/Truth/

Truth-of cluster of semantic concepts is valuable for many purposes, and someone who had one of those purposes in mind might propose that some tokens of definite descriptions be interpreted as Referring expressions. Our arguments show that a defense of the theory of descriptions must be of this broadly theoretical kind, as a thesis subject to a theoretician's discretion. The theory of descriptions cannot be defended as a thesis of ordinary language philosophy, a thesis about what one must accept "in light of what is and is not intelligible to say."

In Part Five I will exercise my theoretician's discretion and propose a semantics for Fred's language that makes no use of the Reference/Truth/Truth-of cluster of semantic concepts, replacing them with a set of epistemic concepts. A philosopher who wishes to dispute my proposals can do so in many ways. But she cannot dispute them by arguing: "*One finds* that when one uses definite descriptions to Refer, one presupposes the existence and the uniqueness of a Referent. *One finds* that it is incoherent to Refer with a definite description while rejecting the Truth of the associated presuppositions. Is that not a defect of Fred's use of 'the big ant,' and as a result, a defect of his use of 'Charley'? And must you not give some account of this fact? You hardly can do so if you operate entirely within a semantics that makes no mention of Reference or Truth." *One finds* nothing of the kind.[3]

10

Trying to Predicate Existence

The argument of Chapter 9 showed that the incoherence one perceives in the conjunction: "There is no nickel in my pocket, and the nickel in my pocket is old," should be understood as ungrammaticality. One is trying to use an anaphor in the face of a commitment to supply no antecedent for it. The theory of descriptions, by contrast, diagnoses the perceived incoherence as a perceived lack of truth-value, which is due to the nonexistence of a Referent under the assumption of the first conjunct. The theory of descriptions misunderstands the incoherence-generating link between first and second conjunct.

My aim in this chapter is to give a supplementary argument for the same conclusion. In outline: one can have a sentence of exactly the same form as "There is no nickel in my pocket, and the nickel in my pocket is old" for which it is not plausible, even on the face of the matter, to explain the incoherence in terms of Reference failure, although the ungrammaticality explanation continues to apply. In order to modify the Reference failure explanation so that it applies (and provided we ignore the arguments of Chapter 9), one must modify the theory of descriptions by replacing Reference, a language-world relation, with an analogue that is not a language-world relation. This is a metaphysically weighty amendment, and the fact that one cannot avoid it if one wishes to stick to a Reference-failure style of explanation, whereas nothing so metaphysically weighty is required if one accepts an ungrammaticality explanation, speaks in favor of doing the latter.

The natural home of this line of thought is in the debate over a classical question of metaphysics: the question whether "existence is a real predicate." Consequently, where one decides to come down within the line of thought just sketched has implications for where one ought to come down

in the debate whether existence is a real predicate. It will be best to fill in at least a little relevant background, so we can locate ourselves comfortably.

Although anticipations of the "real predicate" question can be found earlier, the debate began with a vengeance when Kant claimed that ontological proofs for the existence of God fail because they treat existence as a real predicate, whereas it is not a real predicate. Here is Kant's most interesting formulation of the distinction he is after (he is explaining what goes on when one thinks of a certain quantity of money—a hundred thalers—and then thinks of this money as existing):

> For the object, as it actually exists, is not analytically contained in my concept, but is added to my concept (which is a determination of my state) synthetically; and yet the conceived hundred thalers are not themselves in the least increased through thus acquiring existence outside my concept.
>
> By whatever and by however many predicates we may think a thing—even if we completely determine it—we do not make the least addition to the thing when we further declare that this thing is. Otherwise, it would not be exactly the same thing that exists, but something more than we had thought in the concept; and we could not, therefore, say that the exact object of my concept exists.[1]

Let's change the example so that there can be no doubt that the putative subject of predication is the kind of thing which, if it were real, would be a particular. Quantities of money are a bit iffy in that regard, and we do not want to mix in a problem about distinguishing first-order predication from second-order predication unless that problem is forced upon us. So, suppose I conceive of a horse with certain properties $P_1, \ldots, P_n$. We need to be clear about the scope of "conceive of" here. I do not mean "suppose there is a horse which has $P_1, \ldots, P_n$, and I conceive of it"—this is something I can do without conceiving of the horse as being a horse, or as having $P_1, \ldots, P_n$. I mean "conceive of" to have its longest possible scope in the sentence "Suppose I conceive of a horse which has $P_1, \ldots, P_n$," although there seems to be no way of phrasing such sentences in English that will preclude a shorter-scope reading. One must lapse into a philosopher's pidgin English, and write some such thing as "Suppose I form the concept: *horse that has* $P_1, \ldots, P_n$." Conceiving of a horse which has $P_1, \ldots, P_n$ in this long-scope sense is something that can be done whether or not there are any horses at all.

One point Kant wants to make is that for the horse I am conceiving of

to exist is a matter of the concept I have formed applying to some real individual, and is not a matter of my complicating my concept with the further property of existence; it is not a matter of my forming the concept: *horse that has $P_1, \ldots, P_n$ and existence*. The scope distinction we just noted is helpful here too: when you have a sentence of the form "I am conceiving of a horse which is F and which is G," the scope of "conceives of" can extend to the end of the sentence, so that the meaning is "I am forming the concept: *horse which is F and also G,*" or the scope can stop just short of "and which is G," so that the meaning is "I am forming the concept: *horse which is F,* and this concept applies to something G." But this scope distinction is insensitive to the choice of predicate G. For the horse I am conceiving of to be in the barn is a matter of my concept applying to something that is in the barn, not a matter of my forming the concept: *horse that has $P_1, \ldots, P_n$ and being-in-the-barn*. Kant recognized the scope distinction when "G" is schematic for "existent," and it is clear that he recognized that "I am conceiving of a horse which is F and existent," with short scope for "conceiving of," does not imply the same sentence with long scope for "conceiving of." It is less clear that Kant recognized that the long-scope reading does not imply the short-scope reading, but one imagines he did recognize it (noticing the scope distinction is nine-tenths of the game, and he had noticed it).

It would seem that ontological arguers blur this scope distinction, and therefore do not appreciate that at some juncture they are invalidly inferring the long-scope "I am conceiving of a thing which is $D_1, D_2, \ldots, D_n$ and which moreover is existent" from the short-scope "I am conceiving of a thing which is: $D_1, D_2, \ldots, D_n$, existent." (The predicates $D_1, D_2, \ldots, D_n$ are suitable "defining" attributes of a deity.) There is no doubt that this neatly fingers the source of invalidity in some ontological proofs, if we take the position that they are coherent, intelligible, but invalid. Since Kant might have wanted to make exactly that point, this is evidence that we are reading Kant accurately (though we are supplementing his arsenal of analytical concepts with some fancy contemporary ones). But since the point does not turn on any peculiarity of existence, as opposed to being-in-the-barn or any other quite normal predicate, maybe we are missing some element of his meaning. Here is a likely candidate for the further thing he may mean, again phrased in terms of horses:

> (KANT): Whenever someone says (judges) "I am conceiving of a horse which is F and exists," "conceiving of" *must have* short scope.

> *It is not possible* to mean by that sentence "I am forming the concept: *horse which is F and exists*." This generalizes. Nothing special about horses.

It is important to understand the difference between (KANT) and the very different position that existence is implied by such mundane properties as being a horse, so that forming the concept of *horse which is F* and forming the concept of *horse which is F and exists* are logically equivalent conceptual acts. Kant's remark that a hundred real thalers do not contain the least coin more than a hundred possible thalers seems to point toward this latter doctrine. But his explanation of that remark is the passage quoted above; there is no independent evidence that he believed one can say "I am conceiving of a horse which is F and exists" and mean that one is forming the concept of *horse which is F and exists*.

The principle (KANT) can play no role in showing that ontological proofs are coherent but invalid, though it might be adduced to show that they are not even coherent. The objection would be that an ontological arguer must claim to have established the truth of a premise of the form "I am conceiving of a thing which is D_1, D_2, . . . , D_n, existent." But according to (KANT) no such conceptualizing is even possible. So the ontological arguer represents herself as performing a conceptualization she cannot have performed. Kant's text is too fuzzy to enable one to tell whether he meant to make this point, rather than the point that ontological proofs are coherent but invalid because of a scope fallacy. So let us just assume he meant to make the point that existence is not even a possible conceptual content. This is close to saying that existence is not a property, a revision of Kant's thesis popular among later writers.

After philosophy took its linguistic turn in the twentieth century, various philosophers set about showing that the predicate "exists" plays a different "grammatical role" from that played by other predicates capable of occupying the same position in surface grammar. Two lines of thought stand out:

(I) In the surface grammar, "exists" looks to be a predicate, but actually it is a fragment of a quantifier;
and
(II) In the surface grammar, "exists" looks to be a first-order predicate, but actually it is a second-order predicate. It is used to ascribe a property to a property, or to a kind, or to some other "higher-order" subject.

If either (I) or (II) is acceptable, then we can give one or the other of these arguments:

(Argument I). Suppose a person coherently conceives of a thing which is A, B, C, . . . , and exists. Now "exists" is not a predicate at all, despite surface appearances. It is not used the way predicates are used—to express or signify a property which one might ascribe to or deny of objects. So the person who conceives of a thing which is A, B, C, . . . , and exists is not forming the concept "thing which is A, B, C, . . . and existent," since this last formulation smuggles in the idea that there is a property "being existent" on all fours with "being A," and the rest. Generalizing, "exists" cannot be used, meaningfully, to express a conceptual content.

(Argument II). Suppose a person coherently conceives of a thing which is A, B, C, . . . , and exists. And assume each of "A," "B," "C," . . . is a first-order predicate, a predicate of things not themselves predicable (for example, horses). Now "exists" is a predicate of predicables, despite surface appearances. Since "exists" expresses a second-order property, if the person were including existence in her concept alongside a bunch of first-order properties, she would be making a category mistake and would not be forming a coherent concept. But by hypothesis the person is forming a coherent concept. So the person is not using "exists" to express a conceptual content, regardless of superficial grammatical appearances. Generalizing, "exists" never is meaningfully used to express a conceptual content when the remainder of the concept applies to first-order things, things not themselves predicable.

Both of these arguments move from some version of the thesis that the verb "exists" is not really used as a predicate despite superficial appearances to a slightly heavier metaphysical conclusion, a conclusion one could marshal in order to rebut an ontological arguer. One or the other of the arguments is acceptable—really neat, in fact—if either (I) or (II) is acceptable. Our focus will be on the way linguistic philosophers have tried to defend (I) or (II). Let's explore a defense of (I)—or maybe it is a defense of (II)—that takes the theory of descriptions as a premise.

The defense of (I) I have in mind is due to Richard Gale.[2] Like all arguments for (I) or for (II) proposed by mortals, it only takes into account a proper subclass of existential statements. One needs to add "And this is pretty much how it always goes." It is assumed that once the scales have fallen from one's eyes, one will find a way to see other uses of "exists" as nonpredicative. Here is the argument:

Let us say that a sentence of the (surface) form "The F is G" is being put

to its "standard use" when the speaker purports to use the definite description "the F" to Refer to some entity in the world, to which entity the speaker purports to ascribe the property "being F." When a sentence of the form "The F is G" is put to its standard use, the speaker presupposes that a single appropriate Referent is to be found in the world. We will take the classical view that what this amounts to is that the speaker presupposes that an F exists, and only one F exists (ignoring martini problems).[3] Should that presupposition not be satisfied, the speaker's speech-act misfires, and the speaker has made neither a true nor a false statement. So far this is just a summary of the theory of descriptions.

Now consider a fight promoter who brags that

(1) The fighter who can beat Tony doesn't exist.

If (1) were being put to its standard use, it would lack truth-value unless the presupposition

(2) There is a fighter who can beat Tony

were true. But if (1) is true, (2) isn't. So if (1) is being put to its standard use, it is untrue. But we can imagine that it is true. So (1) is not being put to its standard use.

If (1) is true it cannot have its standard use. But the promoter might be wrong, (1) might be false. If (1) is false, we cannot argue that it must not have its standard use, since (2), the presupposition of a standard use of (1), would then be true. But it is unreasonable to posit that (1) changes grammatical form according as it is true or false, so we ought not construe it as a singular predication when it is false, either. A plausible suggestion is that "the (. . .) doesn't exist" combines a quantifier and negation, so that (1) "really says":

(1*) It is not the case that there is a fighter who can beat Tony.

The *apparent* predicate "exist" in (1) is not *real*. Really the word "exist" is a fragment of a quantifier, the quantifier "the (. . .) exist(s)."

Analogous reasoning leads to the same result for some positive existentials. Suppose another promoter takes issue with (1), and says:

(3) The fighter who can beat Tony sure does exist. You can see him any day you want right here in my gym.

Assume this is false. The first guy's fighter really is the best around. If (3) were being put to its standard use it would not be false. Reason: the

presupposition "There is a fighter who can beat Tony" is false, leaving (3) neither true nor false. (3), like (1), is a singular predication only in the surface grammar.

Finally, we assert that (1) and (3) are typical of the breed. Sentences that are singular predications in the surface grammar are not real predications, and "exists" is not a real predicate. Against this line of thought one can try to make a singular existential—say, a negative one—be a real predication by semantic ascent, as in

(4) This computer lacks existence.

Is not (4) false? The reply is that to think so reflects nothing but bad metaphysics. Nobody learns "what existence is" by finding some in a desk drawer. If we think there is such a property it is because we believe, as a matter of common sense philosophy, that verbs really stand for properties, or we believe something very close to that. We think there is a property of existence to be spoken of because we have not worked through the line of thought that shows precisely that "exists" does not play a predicative, property-expressing role in statements employing it as a predicate in the surface grammar. Since there is no "property of existence," (4) is not false at all, but is without truth-value, for failure of the term "existence" to refer. Since "lacks" is a two-place relational predicate, the nominative "existence" is one of the sentence's two subject expressions. If the theory of descriptions is correct, surely it is also correct to hold that when even one of a relational sentence's several singular terms fails to Refer, the sentence is without truth-value.

How might we tweak this line of thought into an argument for (II)—the proposition that "exists" does express a property, that existentials are genuine predications, but the property ascribed is a property of properties, or a property of kinds, or a property of "roles," or a property of some other category of "exemplifiable"? Answer: choose different existential sentences as paradigmatic examples. For example, suppose a realtor says to you:

(5) The house you've been looking for exists! It is on the North Side.

It is not plausible to think of the "real" grammar of (5) as combining a first-order quantifier with a predicate upon which it operates. The realtor is not implying that some house on the North Side exemplifies the property "being a house you have been looking for." A fortiori, the realtor is not implying that some house on the North Side uniquely exemplifies

"being a house you have been looking for." You have not been looking for that house at all; indeed you may never have looked for any of the houses on the North Side. Plausibly, the "real" grammar of (5) is laid bare in:

(5*) The *kind* of house you've been looking for has an exemplar. It (the exemplar) is on the North Side.

The idea is that there is a kind of house you want, in the sense that you want a house fitting those specifications. You have been looking around for such a house, and there is one on the North Side, or so the realtor claims. Had the realtor glumly told you

(6) The house you've been looking for does not exist,

it would be plausible to understand this as meaning

(6*) The kind of house you've been looking for has no exemplar.

A consequence of this account is that when the realtor utters (5), she presupposes that

(7) There is a kind of house you've been looking for.

And when she utters (6) she likewise presupposes (7). Were the realtor to say "The house you've been looking for doesn't exist, although there is no kind of house you've been looking for (you have not been looking for any kind of house)," she would have "contradicted herself," in just the way a person who says "the nickel in my pocket is old, although there is no nickel in my pocket" contradicts herself. Therefore we are deprived of the Gale-style argument against treating "exists" as a genuine predicate. But that is how it should be. In these realtor's uses, "exists" does play a property-ascribing role. But the ascribed property is "having an exemplar," a property of kinds, not a property of houses (or deities). The realtor's definite description "the house you've been looking for" Refers, but what it Refers to is a second-order object, a kind of thing. (At least this is where we wind up if we unflinchingly apply the theory of descriptions.)

The case for claim (II) above, as against claim (I) above, depends on whether the realtor's use of "exists" makes a better paradigm than the fight promoter's use. Since both the fight promoter's use and the realtor's use are fairly common, and since they point toward different pictures of the semantics of existentials, it would be more reasonable to endorse both (I) and (II) according to circumstances than to opt for either alone as the single correct picture. But someone might argue that there is a class of ex-

istential (surface) predications for which neither (I) nor (II) is quite the right account. That brings us to the point of this chapter.

The existentials I have in mind are (surface) predications of the forms "The F exists" and "The F does not exist" where the description "the F" contains intentional language, or certain other non-existence-entailing language. Typically, a person who utters such an existential is thereby "committed to a presupposition" of the form "There is an F," which is not plausibly construed as meaning "An F exists in the real world."

Suppose, for example, that from time to time Susan daydreams about an old man, although she does not know how these daydreams started. She does not know whether she invented the old man during an especially boring lecture, or read about him, and if she read about him, whether she was reading a novel or a newspaper. Or maybe somebody else made him up and told her, or read about him and told her. And so forth. These epistemic features of the case are logically inessential, but they make the telling easier by suppressing some "pragmatic" noise at certain points. Now consider the existential sentences:

(8) The old man Susan daydreams about exists.
(9) The old man Susan daydreams about does not exist.

Suppose the fact is that Susan saw this man on the street a few years ago, but has forgotten. And suppose the old man is still alive, living happily in Arizona. Then (8) is true and (9) false. Alternatively, suppose Susan invented the old man, but has forgotten that she did. Then (9) is true and (8) false.

Now suppose you say:

(10) There is no old man Susan daydreams about, and the old man Susan daydreams about does not exist.

To affirm (10) is to "contradict" oneself; it is to say something incoherent, just as surely as one "contradicts" oneself or says something incoherent by affirming:

(10Nickel) There is no nickel in my pocket, and the nickel in my pocket is old.

In Chapter 9 we attributed the "contradiction" in the likes of (10Nickel) to a grammatical breakdown, resulting from trying to use an anaphor while declining to accept any sentence containing a suitable antecedent—as one is obliged to do once one endorses the first conjunct of

(10Nickel). The "contradiction" one faces trying to affirm (10) has the same explanation. But a proponent of the theory of descriptions interprets the source of "contradiction" in (10Nickel) differently. She wishes to explain it as resulting from one's failure to Refer successfully with the expression "the nickel in my pocket," with the consequence that "The nickel in my pocket is old" lacks truth-value. Fair's fair, so she ought to give the same account for the "contradiction" (10). That would amount to claiming that the first conjunct of (10) is a negative existential to the effect that there exists no object in the world which has the property "being an old man Susan daydreams about."

Of course the first conjunct of (10) does not mean that; and if it did there would be no "contradiction." Nothing the least contradictory about

(10Fine) There exists no object in the world which has the property "being an old man Susan daydreams about," and the old man Susan daydreams about does not exist.

In fact the first conjunct of (10) is logically particular in form (to advert to old jargon), but it is not existential at all.[4] This does not affect the missing-anaphoric-antecedent diagnosis of the "contradiction" found in (10), since that diagnosis makes no assumptions about the semantics of E-sentences. They might be interpreted as objectual quantifications over a domain of existent things—the denizens of a real world—or they might be interpreted as objectual quantifications over a pumped-up domain including lots of nonexistent, merely "intentional," objects; or they might be interpreted as substitutional quantifications. Sentences of the form "There is an F" would be genuine "existential" generalizations only in the first case (of course it is common parlance in logic to call E-sentences "existential generalizations" in virtue of their form and their inferential role, as I have done until now).

It is obvious that a person who says "The old man Susan daydreams about does not exist," in any of the circumstances we have discussed, is "referring to the old man Susan daydreams about" in the conventional meaning of those words. The person is talking about the old man Susan daydreams about, alluding to the old man Susan daydreams about, making mention of the old man Susan daydreams about. The conventional, nonphilosophical meaning of "refer" does not require that what one refers to has to exist; "person X just used expression E to refer to Y" is nonextensional at the position of "Y," and does not imply that Y exists.

But what if a diehard philosopher digs in her heels and insists on ex-

plaining the contradiction in (10) in terms of the theory of descriptions, or something as near the theory of descriptions as she can have—how might that account go? The intuitive data needing explaining boil down to this: when you utter (9), you must not deny the logically particular generalization "There is an old man Susan daydreams about," and you must not affirm that there is in addition some other old man Susan daydreams about. Here's a sample theory.

We posit a relation Reference* obtaining between certain terms in a language and a domain of objects D, the elements of which are "intentional objects" understood on analogy with characters in a story. These "thought-characters" are drawn from the thought of all the people you and I could have reason to discuss. As it happens, one of the characters in Susan's story is an old man she is forever daydreaming about. You and I can use terms that Refer* to items in this domain; one such term is the definite description "the old man Susan daydreams about." When one of us uses the expression "the old man Susan daydreams about" to Refer* to an element of D, there must be such an element (an element fitting the description), on pain of the speech-act misfiring. This implies that when you say something of the form "The old man Susan daydreams about is (isn't) F," you presuppose that a suitable element can be found in the domain D. We maintain that in English the fact that such an element can be found in D is expressed by the generalization "There is an old man Susan daydreams about." That generalization exhibits the inferential behavior of an existential generalization, but should be understood semantically as an objectual quantification over D.

One predicate you might apply to "the old man Susan daydreams about" is "exists," or if you prefer, "doesn't exist." We need some account of what it is for an element of D to satisfy these predicates. Here's a quick one. Like characters in a story, elements of D can be based on real things or not. If they are, they satisfy "exists"; if not, they don't.

Now suppose you utter (9); you say "The old man Susan daydreams about does not exist." You are committed to the presupposition that there is an old man Susan daydreams about. This, we claim, is because you are attempting to Refer* to whichever element of D has the property "being an old man Susan daydreams about," and so it had better be there, in D. End of theory.[5]

If we aim to diagnose the "contradiction" in (10) in terms of something *like* Reference failure, Reference* failure is the kind of thing we will be driven to posit. We must, after all, explain why the speaker presupposes a

logically particular sentence that lacks existential import. Reference* is not a language-world relation, though it is a language-something relation, where the something is a domain of intentional objects, possibly real, possibly not.

So, here is the situation we have: an utterer of "there is no nickel in my pocket, and the nickel in my pocket is old" says something incoherent, in some sense of "incoherent," and an utterer of "there is no old man Susan daydreams about, and the old man Susan daydreams about does not exist" also says something incoherent, in some sense of "incoherent." We have a diagnosis of the first incoherence that extends smoothly to the second—the utterer is declining to put on the table a suitable anaphoric antecedent for an anaphor. This diagnosis makes no theoretical use of the theory of descriptions. A philosopher who wishes to reject our explanation and revert to the theory of decriptions, explaining the incoherence of "there is no nickel in my pocket, and the nickel in my pocket is old" as a matter of truth-value-less-ness, ought to give a parallel explanation for the incoherence of "there is no old man Susan daydreams about, and the old man Susan daydreams about does not exist," or else say why these two very similar instances of "incoherence" deserve radically different treatment. Assume that this philosopher chooses to apply "the theory of descriptions," or the nearest possible thing to it, in the case of "there is no old man Susan daydreams about, and the old man Susan daydreams about does not exist." This will require positing a Reference-like relation which connects language with intentional objects.

I will make no effort to recapitulate the centuries of metaphysical bewilderment engendered by doctrines positing "domains of possibly unreal intentional objects." If the reader has not suffered this bewilderment, here is a good way to start suffering. Imagine that you have become persuaded that some foreign agents have moved into your spare room, thinking to brainwash you into complicity. You have not settled on an exact number of foreign agents, since that is the sort of thing one needs to examine carefully. As a consequence of this belief of yours, how many intentional objects are injected into the domain of "possibly unreal things you believe are in your spare room"? If the answer is "no particular number," then what sort of thing is this "domain"? Can it be a set? (Do not go off talking about "fuzzy sets"—there is no issue here about *well-defined* entities being to some extent members of a set and to some extent not.)

In Chapter 1, we briefly considered the option of interpreting a confused person as using terms that Refer to what I am now calling "inten-

tional objects" (this was Option 4 of the five semantic options we surveyed). Anyone who elects that option must confront the kind of puzzle just described. I believe certain perfectly respectable philosophical goals might be achieved most efficiently by adopting Option 4—and taking on the obligation of sorting out the puzzles one way or another—though in this book I will adopt a form of Option 5. But the goal of defending the theory of descriptions, in the face of what seems a better explanation for the intuitions in which it is grounded, is not a weighty enough philosophical goal. We ought to prefer the metaphysically lightweight alternative of a syntactic explanation for both the "incoherence" of "there is no nickel in my pocket, and the nickel in my pocket is old," and the "incoherence" of "there is no old man Susan daydreams about, and the old man Susan daydreams about does not exist." In both cases what we have is the incoherence of anaphoric-antecedent-less-ness.[6]

This argument gives us a reason for rejecting the theory of descriptions even in the case of the "incoherence" of uttering "there is no nickel in my pocket, and the nickel in my pocket is old." But that is a paradigm case of the alleged "incoherence of predicating when there is no suitable Referent" that is the driving intuition of the theory of descriptions in Strawson's form. If the theory of descriptions should not be adopted to explain what is going on in this case, when ever should it be adopted?

Obviously this argument is supplementary to the argument of Chapter 9. The bottom line for our purposes remains the same: we cannot argue that when Fred says something of the form "Charley is F," and we, in a judgmental rather than a humoring temper, decline to ascribe a truth-value to what Fred has said, this is because we know Fred does not use the term "Charley" to Refer and Refer uniquely to some real ant. The right explanation for our disinclination to truth-value Fred's statement is given by another line of thought entirely: the calibration argument of Chapter 7.

V

A LOGIC FOR CONFUSION

11

Explicating

The upshot of Part Four was that we should not look to the Reference/Truth/Truth-of cluster of semantic concepts if we wish to find a semantics that fills in the details of a confusion-attributer's charitable semantic proposal.[1] It follows that we cannot employ a classical validity criterion: validity understood as guaranteed truth-preservation. We need to identify a different conception of validity; one that does not involve the Reference/Truth/Truth-of cluster at all.

Let us loosely define a class of arguments as "declarative arguments." The definition is loose because it is partly by paradigm: a conjunction elimination containing only declarative sentences, set forth in a learned article in the sciences, is a piece of declarative argumentation. So is an inference to the best explanation, set forth in the same learned article. So is a universal instantiation, so is a statistical inference. Everyone knows how to continue this paradigm: these are the arguments one studies in "Intro Logic," "Intro Stat," and "A Seminar in Tense Logic." They are not the arguments one (primarily) studies in "A Seminar in the Logic of Questions" or "A Seminar in the Logic of Imperatives." If some of the declarative sentences in an argument comprised entirely of declaratives are being used to issue commands or pose questions, so that they "really are" imperatives or interrogatives, the argument does not count as "declarative."

This definition by paradigm is accompanied by a functional definition: when an argument is put forth *to prove that things are a certain way,* or *to give an explanation why things are a certain way,* the argument is declarative. Everyone knows how to identify members of this functionally defined class, though philosophical differences may lead to quarrels in a few cases. For example, differing accounts might be given of arguments that "would be" declarative arguments if some arguer actually endorsed them, but which in fact are merely displayed, perhaps as illustrations in a logic text-

book. Someone might think they are not arguments at all but structures which could be used as arguments; someone else might think they are being used to pretend to prove things or explain things; there are still other options. Since the functional definition of a declarative argument does not settle these questions, it, too, is loose. The concept of a declarative argument is epistemic, because the concepts "prove" and "give an explanation" are epistemic.

The classical concept of validity is part of a philosophical theory about what the main function of declarative arguments is: to prove that things are a certain way, or explain why things are a certain way. To put it differently, this philosophical theory holds that the characteristics of declarative arguments we just used to define the class in fact get at the fundamental point and purpose of declarative arguments (whereas in general defining characteristics need only get at *salient* necessary and sufficient characteristics). This is essentialist talk, and wrongheaded. But a weakened, nonessentialist version of the theory is correct: the concept of "following-from" can be *explicated in terms of* the concepts "prove" and "explain." This thesis is compatible with holding that the concept of "following-from" also can be explicated in terms of other families of concepts.

To speak of "explication" here is to imply that the concept of "following-from" is a confused concept, like Fred's concept of "Charley," or Marvin's concept of "cooler than."[2] Suppose we taught Fred to identify Ant A by a microscopic mark, and we taught him to use the name "Ant A" all the time as a name for that ant, never using the name "Charley" at all. Fred would have an unconfused language, though an impoverished one. He would have no linguistic means for responding to the presence of Ant B. If we allowed him to use the name "Charley" again, provided he used it as an exact synonym of "Ant A," we would have taught Fred to adopt what I am calling an "explication": in this case an explication of his concept of "Charley." An explicated concept is unconfused, but conceptually impoverished.[3]

If we were careful not to read too much into it, we could say: "In coming to use an explicated 'Charley' concept, Fred has adopted the philosophy that only Ant A matters when it comes to making judgments about the ant colony." That is how it is when we explicate the concept "follows from" as classical validity. There are other "things one means" when one says a conclusion follows from premises, but one has run these things together with each other, and with classical validity. The prephilosophical concept is rich, but it is too rich. Once we explicate it as classical validity, it

is (relatively) impoverished, but usable without confusion. If we are careful not to read too much into it, we can say: "In coming to use this explicated concept of following-from, we have adopted the philosophy that only proving things true and explaining why things are true matter when it comes to evaluating declarative arguments." In this special sense of "adopt a philosophy," it is a sensible policy to adopt the traditional philosophy that the main function of declarative arguments is to prove and explain.

But in this special sense of "adopt a philosophy," it may also be a sensible policy to adopt a contrary philosophy, and be ready to shift off from one to the other as needed. Fred would be well advised to adopt the philosophy that only Ant B matters as a standing alternative to his philosophy that only Ant A matters. We could teach him to do that by asking him to go back to square one again, using the name "Charley" as before, and then teaching him to use the name "Ant B" as we use it, dropping his use of "Charley." Then we would allow him to reintroduce the name "Charley," but as a synonym for the name "Ant B." This elaborate process of education, added on to our earlier elaborate process of education, would leave Fred in a position where he had two uses for the name "Charley," each corresponding to an explication of his original, confused, concept. But once he was aware that both explications exist, he would not be led to think that whichever philosophy he had adopted at the moment was best. He would know that each had a coequal rival.[4]

The classical concept of validity is an explication of a confused prephilosophical concept of following-from. Or so I assert. To support this assertion, I need to describe an alternative explication. Let us call this hypothetical alternative "following-from as X-preservation," where X is some (hypothetical) property of sentences of very great importance to the purposes for which we make declarative arguments. Indeed, X must be of such fundamental importance that we can comfortably adopt the philosophy that X is what matters when evaluating declarative arguments—though recognizing that there also is the contrary classical philosophy, itself entirely fit to adopt.

We can find an X that fits the bill, or rather, two related Xs that fit the bill, if we turn our attention to a certain type of practical deliberation sometimes called instrumental practical reasoning, the chief aim of which is to identify what steps ought to be taken to achieve some goal. To put it roughly, the properties X will be properties of sentences such that when a sentence has a property X, then believing the sentence will contribute to the believer's achieving some goal he or she seeks to achieve. It will take a

while to flesh this idea out. Having fleshed it out, we will give an explication of following-from for declarative arguments as guaranteed X-preservation. Like the classical concept of validity which explicates following-from as guaranteed truth-preservation, indeed like all explications, this explication will leave us with an impoverished concept, but an unconfused one.

If we adopt the suggested explication, we will be treating the Xs as the properties which valid arguments should preserve *whether or not they occur in practical reasoning,* even though the significance of the Xs is that when a sentence possesses an X this marks it as belief-worthy for an actor. We will be making practical effectiveness the measure of worth in conclusions, and therefore the measure of what reasoning ought to preserve. Similarly, the classical point of view is that truth is the property that valid declarative arguments should preserve whether or not they occur in "theoretical" reasoning. A philosopher who holds the classical view insists that when a declarative argument occurs in the course of practical deliberation, the reasoner should be aiming at getting a true conclusion of that declarative argument (though the practical reasoning as a whole might be nondeclarative—perhaps issuing in an imperatival conclusion—and truth might not be the goal of this nondeclarative reasoning).

Because we will know that both options for explicating the concept of "following-from" exist, we will not be as poorly off as a philosopher who accepted one or the other without suspecting the existence of its alternative. For example, we will be able to take up the nonclassical option when our aim is to describe a criterion of validity that ratifies as valid those of Fred's arguments we wish to endorse in a spirit of inferential charity. But we will be able to shift to a classical point of view on validity for other purposes, if we so choose. But that is not the same as having a one-size-fits-all criterion of validity with no coequal alternative in sight. It does not enable us to say absolutely what it is for a conclusion to "follow-from" premises. I would be happier with a one-size-fits-all criterion. But the world is not here to make philosophers happy.

12

Good Advice

When one first thinks about ontological confusion, it is natural and intuitively plausible to talk in terms of perspectival truth. One wants to say: "what the confused person thinks may be true from one perspective but false from another perspective; or it may be true from both perspectives, or false from both." Perspectival truth must replace truth *simpliciter* when one evaluates a confused belief. To avoid assimilating confusion to other sorts of "ambiguous" thought—thought involving equivocal expressions, like "bank," for example—one is inclined to add, "but the confused person cannot distinguish the two (or more) perspectives."

This line of thought needs work. For one thing, it seems circular to explain "cannot distinguish" by saying "cannot distinguish the two perspectives." And a perspective must be somebody's perspective, or (stretching the idea to near breaking) a perspective that could be taken by someone. But our intuitive formulation mentions nobody whose perspective is the operative one, and does not hint at what "taking" a perspective might be. Let's try to make enough improvements in the intuitive line of thought to give us a workable semantic conception, guided by the spirit of the original.

Let's introduce two people with different "points of view" on Fred's ant colony, each point of view expressed in the same words Fred uses, including the word "Charley." But one of the two people will use the word "Charley" as a name for Ant A, while the other uses it as a name for Ant B. Our idea, when we finally give its details, will be that a sentence in Fred's ant-colony language can be assigned one "opinion value" for each of the two people, a value entirely determined by that person's opinion. Neither of the opinions on a given sentence can be "right" when the sentence is used the way Fred uses it (as contrasted with the way one or another of the

two opinion-makers uses it), because Fred's ant-colony sentences cannot be truth-valued.

But one or the other or both opinions may carry weight for Fred as practical advice, even if the sentence in which the advice is formulated is not truth-valuable as a sentence of Fred's language. This is because when Fred is dealing with his ant colony, he can *profit from the fact* that Ant A has, or lacks, a certain property even though he cannot represent that fact in language or in thought. Likewise, he can profit from the fact that Ant B has a property even though he cannot represent that fact. This is one of the main "metaphysical" differences between practice and reason: if some stuff is really there, you can as a matter of fact trip over it or as a matter of fact open a package with it whether or not you even can conceive that it is that stuff you are knocking against or manipulating. Of course if an agent is profiting from a fact she cannot represent, then she cannot represent the further fact that she is so profiting, but perhaps others can. These others might be more learned, less psychotic, more psychotic, situated on the leeward side of a vast alteration in the sciences—or they might have noticed a second big ant.

We will grab this idea and run with it. The two people who have the two "perspectives" on Fred's ant colony will be able to make judgments about Ant A, or Ant B, unconfusedly, and they will express their judgments in unconfused variants of Fred's language: one of them using "Charley" to name Ant A, the other using "Charley" to name Ant B. Suppose the person who uses "Charley" to mean Ant A observes Ant A nibbling a leaf. This person says to Fred "Charley is nibbling a leaf." Suppose further that the person who uses "Charley" to mean Ant B observes Ant B nibbling a leaf. This person says to Fred "Charley is nibbling a leaf." We know that if Fred goes to "find Charley" in order to obtain a nibbled leaf, he will profit from both reported facts, although he cannot represent either of them in thought. Fred will get his nibbled leaf, and we know why. (The "we" here is intended to be both "we in the story" and "we, the philosophical appraisers of the story.") If Ant A were nibbling a leaf but Ant B were not, the account of how Fred would "profit from" the fact that Ant A is nibbling a leaf is slightly more intricate, though only slightly. If you want the general idea, ask yourself how many creatures around you right now are nibbling leaves, and then ask yourself whether Fred is more likely than you to find some nibbled leaf even if only one big ant is nibbling. (If you read that last sentence sitting in your yard, go indoors and read it again.)

Some preliminaries. Suppose S is a sentence of Fred's ant-colony lan-

guage. Suppose S1 and S2 are that same sentence, but with "Charley" used as it is used by each of the perspective-definers. (Assume "Charley" is the only "confused" expression we need to worry about.) Call the perspective-definers Sam and Sal. Sam uses "Charley" as a name for Ant A, Sal uses it as a name for Ant B. Sam's meaning-variant of S is S1, Sal's meaning-variant of S is S2.

Suppose Sam announces "S1." For concreteness, let S1 be "Charley is nibbling a leaf at this very moment." Under the terms of the very simple story we are telling, it is clear what it would be for Sam's remark to be true: Ant A would have to be nibbling a leaf. Suppose this fact obtains. Sam does not make his announcement carelessly. He has made scrupulous inquiries, of a sort appropriate to confirming S1, and he has a fund of general knowledge and a set of intellectual skills that equip him well to make such inquiries. Let's summarize that by saying Sam's announcement is "authoritative," pushing the ordinary meaning of "authoritative" a bit. In particular, we do not require that an authoritative statement be true, although Sam's happens to be. We are not really interested in the truth of Sam's statement, for reasons which will emerge shortly. (Those reasons also will force us to be more careful about how we define "authority.")

Authoritative statements about what ants are doing call for empirical methods of confirmation, so Sam has applied such methods. He has made observations, run experiments as need be, made predictions, seen whether they held up, and so forth. His predictions have held up. There is some explanation for this. Here the explanation is simple: Ant A is indeed nibbling a leaf, and Sam's "tests" of the sentence S1 have picked up on this fact, and have picked up on a host of closely related facts (related logically and causally), such as the fact that some ant is nibbling something, and the fact that there is freshly nibbled leaf to be found. Therefore, Sam's statement that S1 is an indicator that some facts exist that Fred can profit from if he acts on the belief that S—that is, the belief that "Charley is nibbling a leaf at the moment"—with these words taking the meaning they take in *Fred's* language.

Here is what I mean by "can profit from." The facts Sam has picked up on are there, and Sam has said "S1" because he has picked up on them. Given how we have constructed the relationship between Fred's sentence S and Sam's sentence S1, if Fred accepts (as true) the sentence S, and acts on this belief "in his usual fashion," Fred's prospects of achieving the goals of his action will be enhanced. The idea of Fred acting on a belief "in his usual fashion" deserves a comment.

We imagine Fred to be typical of people in the way his beliefs play a (causal) role in his actions. If he wants a drink of water, and he believes there is a water fountain around the corner, this belief sometimes, perhaps often, plays some causative role in his going around the corner, if indeed that is what he does. It is not for philosophers to figure out how that process works, how beliefs play a role in behavior. It is for psychologists to figure out. If they come back and tell us that beliefs apparently do not play any role in behavior, despite appearances to the contrary, we will need to come back to the drawing board. We'll assume that common sense gets it roughly right at least in this respect: given the role Fred's beliefs play in his behavior, when he wants a drink, it will be to his advantage to have the belief that there is a water fountain around the corner if there is a water fountain around the corner, whereas if there isn't one, having the belief will be to his disadvantage. If Fred were cognitively wired in a bizarre way, perhaps given to acting in as self-defeating a way as one can imagine at every turn, this might be due to his beliefs playing some role in behavior atypical of people. Then, believing the sentence S might bode ill for his goal-achievement. This sort of "backward-wired-person" story makes at least superficial sense, so we stipulate that Fred's beliefs serve him "normally."

When the sentence S is such that if Fred were to act on the belief that S, there would be some facts he would profit from, let's call the sentence S, and also Fred's belief in S, "profitable for Fred." We'll drop the qualifier "for Fred" most of the time. When the sentence S is such that if Fred were to act on the belief that S, there would be some facts that would "cost" him, as measured by failure to achieve his intended goals, let's call the sentence S, and also Fred's belief in S, "costly for Fred" (or just "costly"). Now, if Sam, the authority, has said (affirmed) "S1," and said it as a result of careful inquiry, one expects the sentence S to be profitable for Fred. The same goes if Sal has said "S2" with the symmetric conditions satisfied. And if Sam (Sal) *denies* S1 (S2), one expects the sentence S to be costly for Fred. In our simple example of confusion, it may seem that this is because when Sam says "S1" after careful inquiry, probably S1 is true. So when Fred acts on the belief that S, he is likely to profit from the fact *by virtue of which S1 is true*. That diagnosis is fine here, but insufficiently general. Another toy example will help make clear why.

In the good old days there were two scientific communities, the Predecessors and the Successors. The Successors came along right after the Predecessors and made a lot of advances.

For example, the Predecessors applied the relational predicate "x is

cooler than y" by feeling a pair of objects with their hands, or holding the objects against their faces, or walking on them in the event they were barefoot (just like Marvin). Similarly for "x is warmer than y" and "x is neither warmer not cooler than y." The Predecessors possessed serviceable though crude thermometers, but they kept breaking them. Nevertheless, they believed it reasonable to rely on thermometers for supplementary "temperature readings." On one famous occasion they made a pile of hubcaps and dictionaries, to test their conjecture that a pile of stuff left alone in a closet will "reach thermal equilibrium," meaning (to them) that nothing is either warmer or cooler than anything else. (This had seemed likely, following many tests on piles of just hubcaps and piles of just dictionaries.) The Predecessors made startling observations before their thermometer broke, as one would expect.

The Successors, who developed the rudiments of what we call thermodynamics, charged the Predecessors with being unable to distinguish a thing having lower temperature than another thing from a thing having higher specific heat than another thing. When the Predecessors spoke of one thing "being cooler than another" (said the Successors), they were betraying that confusion. The startling results of the abortive mixed-pile-of-stuff experiment were due to just these facts: the stuff was all at the same temperature, as the thermometer showed, but the hubcaps had higher specific heat, and so felt cooler.

Let's make some additions to the story. Suppose the Successors had a little statistical mechanics to go along with their thermodynamics, though the statistical mechanics was what we (in reality) would call "classical." Suppose further that we (in the story) have made essentially the same advances beyond classical statistical mechanics that we (in reality) have made. Except that in the story we have drawn different ontological conclusions from this history than in reality we have drawn. Because the classical statistical models of specific heat were quite significantly different from the quantum statistical models we now accept (in the story and in reality), in the story we take the position that specific heat has been found not to be a real physical property. This would amount to treating the classical statistical models as very tightly conceptually bound to the concept of specific heat, rather than treating specific heat as a concept of phenomenological thermodynamics, associated only contingently with any given statistical model (as in reality we do). In the story, we say that it is perfectly understandable that the Successors believed in specific heat. They had cogent derivations (just as the classical derivations in our real present-day text-

books are cogent), and their experimental focus on high-temperature (and thin-gas) circumstances led them, reasonably, to downplay the occasional poor fit between their models and experimentally measured values—not every bad data point is a theory-falsifier. Indeed, as a general proposition, we think highly of the Successors as scientists.

Now suppose we say of the Predecessors, "They could not distinguish between having lower temperature and having higher specific heat." Have we said something incoherent, or at any rate something incompatible with the rest of our views in science and in the history of science? We could go on to say something incompatible with the rest of our views, by foolishly adding "and these physical states are distinct." But a more natural continuation would be "even though a reasonable explanation of their most famous experiment, the explanation soon given by the Successors, adduced precisely this distinction." What matters is the cogency and reasonableness of the distinction, not whether its terms are real.

Now remember what our philosophical agenda is. We maintain that an attribution of confusion is a taking of a semantic position. We are trying to formulate a semantics that fits the bill, a semantics one can think of as what the confusion-attributer "intends." So the fact that it is coherent for us to attribute confusion to things we do not believe in, but which are such that those who do (or did) believe in them had a reasonable, cogent view of them, suffices to show that our semantics must cover those cases.[1] We accomplish that in the case at hand by letting there be a Sam* and Sal* who stand to the Predecessors' use of "x is cooler than y" (and cognates) the way Sam and Sal stand to Fred's use of "Charley." Sam* and Sal* presumably would be members of our own scientific community, who have learned how to make Successor-style judgments in their required graduate classes on "Classical Successor Thermodynamics." Sam* takes care of interpreting "cooler than," when it occurs in Predecessor sentences, as meaning "has higher specific heat than," whereas Sal* takes care of interpreting "cooler than" as meaning "has lower temperature than." Then Sam* and Sal* give their opinions on sentences of the Predecessors' language. These opinions are "authoritative" not with the implication of "up to date and correct," but only with the implication of "rationally derived according to standards appropriate to the subject-matter." Sam* and Sal*, like Sam and Sal, are to be the ultimate source of the "opinion values" our semantics deploys.

Are the opinions of Sam* and Sal* indicators of profitability and of costliness, like the opinions of Sam and Sal? Surely they are. We (in the story)

do not believe in specific heat, but the Successors did not make up that idea from whole cloth. They used good empirical methods and used them wisely. They made predictions how their experiments would come out, based upon the supposed properties of temperature, specific heat, and other thermodynamic properties and relations. Those predictions turned out to be correct (mostly, nearly). Otherwise the Successors would not have adopted the concepts. Something, or several things, really were going on to make the experiments turn out as predicted. The Successors had picked up on some facts, and, moreover, they had a principled way to pick up on facts that would make their judgments turn out true. Or nearly true. Or nearly true, mostly. It's just that we think they were unable to conceptually represent those facts.

Suppose a speaker of the Predecessor language says, "The hubcaps in that big pile are cooler than the dictionaries." Let this sentence be S*. Suppose Sam* gives his opinion that "The hubcaps in that big pile are cooler than the dictionaries." Let this sentence be S1*. If the speaker of the Predecessor language acts on the sentence, she is putting it to an empirical test, albeit perhaps a crude one. Whatever factor or combination of factors the Successors have misidentified as "specific heat difference" is likely to cooperate with the Predecessor actor, just as it (probably) has cooperated with Sam* when he tested S1*. So when Sam* gives his opinion in favor of S1*, this opinion is an indicator of profitability for the actor who speaks the Predecessors' language.

Given the particular details of this story, it is possible to make more sharply honed predictions about the profitability of acting on Successor-style beliefs: it will be a more profitable thing to do when one is operating near the classical limit than when one is operating in quantum homeland. That means that if we have special information about a certain person, we might judge that the generally authoritative opinions of Sam* and Sal* should be regarded as "less authoritative" relative to this particular person's practical decisions. There is nothing surprising about that: much wise advice concerning the health benefits of taking long walks would not qualify as quite so wise for an audience of people who live amid landmines.

The argument of the last few paragraphs is fairly general, though it applies only to subject matter for which the appropriate methods of inquiry are empirical. For other kinds of subject matter, pure mathematics for example, we cannot claim that the advice given by "authoritative" but mistaken theorists probably reflects the existence of some (poorly conceptualized) "facts of nature" which an actor can profit from if she takes the

advice.[2] Even for nonempirical subject matter, though, it may be possible to argue *in a given case* that we, who are stipulated to be in the know, can coherently attribute confusion to subject #1, where the "objects" we say were confused by subject #1 were held to be separate and distinct by subject #2, who, in turn, we say, had not yet worked out adequate concepts. Quick example:

Augustin-Louis Cauchy could "separate" from one another various limit processes or "modes of convergence" that the earliest analysts could not separate, and that even analysts just before his time (say, circa 1800) could not separate. Karl Weierstrass could separate modes of convergence Cauchy could not separate, having a better command of what we would call quantifier-order considerations.[3] We might well side with Weierstrass, but still adopt Cauchy's framework in order to make the point that the ca. 1800 analysts were more confused than Cauchy, who made a great relative advance. In this case it surely *is* true that someone who relied on Cauchy's conceptual framework in complex analysis would do quite well, so that an affirmative Cauchy-style judgment on a given sentence of analysis is an indicator of profitability. If two different people, each applying Cauchy-style standards, played the roles of a Sam and a Sal vis-a-vis some sentence of circa 1800 analysis, and each affirmed the sentence when it was given the particular spin he put on it, that also would be an indicator of profitability—assuming the sentence is acted on by a person whose goal is effective mathematical work. If such an actor relied on the sentence, she would not stumble into trouble because of the blindness of ca. 1800 analysis to Cauchy's distinctions. Nothing in what I have just said implies that Cauchy's distinctions "got it right," or even implies that they were conceptually adequate by the lights of a Weierstrass.

Summing these points up: sometimes the case can be made that even though a certain person (or community) did not get things right when they distinguished X from Y, the authoritative affirmative opinion of a "Sam" or "Sal" speaking in "X" terms and in "Y" terms is an indicator of profitability. When such a case can be made, it is coherent to say of some other person that she confuses X with Y. When the case cannot be made, it is not coherent to do so. For empirical subject matters, there is a fairly general plausibility argument—the one we gave above—that authoritative affirmative opinion ought always to be an indicator of effectiveness, though no doubt an indicator varying in strength from one case to another, according to just how "cogent" the theoretical framework is within which "Sam" and "Sal" operate. For nonempirical subject matter, there is

no neat general plausibility argument, but there might be one in a given case.

The driving idea behind this entire line of thought is that for a given subject matter there may be people, Sams and Sals, who are "authorities." What is essential about "authorities" is that their opinions carry weight as practical advice, as recommendations that one act on this belief but not on that one. When we, the people who are attributing confusion to some subject, believe in the categories in terms of which the confusion is framed, the concept of authoritativeness could be replaced by the concept of knowledgeability. What we would want from Sam and Sal is a knowledgeable, even expert, application of concepts we share with them. When we, the people who are attributing confusion to some subject, do *not* ourselves believe in the categories in terms of which we frame our attribution, it is inappropriate to speak of "knowledge," thereby leaving the implication that Sam and Sal get things right. Then what is "authoritativeness" if not knowledgeability, or "knowledgeability of a high order"?

Suppose the confusion we wish to attribute involves the terms X and Y. At a minimum, when they classify things as X or as Y, Sam and Sal must conduct their investigation rationally, by a standard of rationality appropriate to the subject matter. To make that just a bit longer than a bumper sticker, look at it this way: there are epistemic rules specifying what conditions justify what opinions. (I am assuming that rational practice is rule-following.) These rules are hard to formulate, much easier to follow, and subject-matter sensitive. An inquiry is conducted rationally when the inquirer obeys the rules. This is only a minimum requirement, because unless some of the rules oblige an inquirer to make "thorough" and "painstaking" investigations, someone could obey the rules but not look very hard for ants, or obey the rules but not test all the objects in a pile for specific heat. Soon we'll see in more detail how this matters. For now, let's just add on a "thoroughness" requirement, as intuitively plausible, giving us this working definition of an "authority":

> An authority bases her judgments on investigations she has conducted rationally, according to a standard of rationality appropriate to the subject matter, and which she has conducted thoroughly.

The term "rationally" is overworked; to put it more colloquially, an authority is really, really good at something. The judgments of someone who is an authority in this sense are indicators of profitability or, when the judgment is negative, of costliness. We have some plausibility argument

for this proposition, at least when the subject matter is accessible to empirical investigation.

Now let's get simple and concrete again. You can't do semantics in the stratosphere. Let's make up a nice story about Fred and Sam and Sal and the ants, putting in enough detail to nail down the idea that when Sam states his considered opinion that S1, or Sal states his considered opinion that S2, these opinions are indicators of the profitability of Fred's sentence S. The considerations that arise for this simple case of confusion illustrate themes that arise more generally, but one should not expect the analogy to be very tight when the subject matter is very dissimilar—for example, when the subject matter is complex analysis and the "items" said to be confused are "modes of convergence." Doing epistemic semantics is bound to be like that. You have to formulate the details for rather small families of quite similar epistemic settings, and then say: "So you see, it goes like that."

The first point to make plain is that these authoritative opinions are indicators *for us* of profitability—or, if they are negative opinions, indicators *for us* of costliness. ("We" being the characters in the story, the people who, in the story, are attributing confusion to Fred.) If Fred knows that Sam and Sal are experts on ant colonies, and knows further that they have been making an especially careful study of his ant colony, he might well accept their advice, or what he takes to be their advice. But Fred has no knowledge of the deviant meanings Sam and Sal attach to his words, so he cannot know what advice *really* is being proffered. So it is uncertain whether Fred can be said to *know* that what Sam (Sal) tells him is a good indicator that acting on the given sentence will promote goal-achievement. The argument that he does not know it would point out that were Fred to be informed that Sam and Sal each has a different ant in mind when he says "Charley," Fred should take his grounds for trusting their opinions to be undercut. (This sort of "undercutting" consideration is an element in Gettier-style arguments, though applicable more broadly.) But it is irrelevant whether this is a good argument, and irrelevant whether Fred knows that what Sam (Sal) tells him is a good indicator that acting on the given sentence will promote goal-achievement. We are trying to describe a criterion of validity for Fred's arguments that can serve as a fleshed-out version of the semantic position we take when we attribute confusion to Fred. It matters that the "opinion values" generated by Sam and Sal be indicators-for-us of profitability-or-costliness-for-Fred. If Fred does not, even cannot, know that the opinions in question have such im-

plications, all that shows is that in semantics ignorance of the law is no excuse: we can go right ahead and claim that Fred's reasoning ought to conform to the validity criterion we endorse.

To make a tractable "story" out of the interactions of Fred, Sam, and Sal, it helps to imagine that we have fibbed to Sam and Sal, so that each actually thinks he is speaking Fred's language with no shifts of meaning. We tell Sam a story about Fred's use of the name "Charley" which embodies Kripke's distinction between the semantic reference and the speaker's reference of a name. We tell Sam that Fred uses the word "Charley" as a name of Ant A, but occasionally errantly calls Ant B "Charley." On these occasions, we say, Ant B is Fred's "speaker's referent," but Ant A remains the semantic referent of "Charley," the referent it has as a matter of its meaning in the language. We tell Sal the symmetrical story, with the roles of Ant A and Ant B reversed. We could tell the story differently, and put Sam and Sal in the know about our plan to use them to define semantic values for a proposed account of validity, but that makes for a much more complicated story, because we would have to add a complicated instruction about how they should ignore this role they are playing when they decide whether to endorse or reject a sentence. That would be like knowing that you really are "the measure of every thing that it is true, if it is true (etc.)," but putting that out of your mind when you "reported facts." It would be difficult for you to stick to your ordinary evidentiary standards. Our epistemologist's intuitions will engage more smoothly as we go along if we do not imagine Sam and Sal burdened in this way.

Sam knows all about what Ant A is up to, and he knows all about what Ant B is up to; ditto for Sal. But in fact we only ask Sam about Ant A, because we formulate all of our questions in terms of "Charley." Using the same technique, we ask Sal only about Ant B. Of course it sometimes happens that in order to figure out something about Ant A, Sam must take into account facts about Ant B, and symmetrically for Sal. But we do not ask Sam or Sal any questions which explicitly distinguish between Ant A and Ant B, because Fred cannot articulate such questions. To the extent that either Sam or Sal takes into account "cross-ant" information, they do it behind the scenes, as it were. Evidently this prevents Sam and Sal from saying things like "Ant A despises Ant B," but we are content to let the "Charley"-talk sublanguage of their languages be expressively incomplete with respect to what they can say in their full ant-colony languages.

It will be helpful to make a special point of sticking in a "thoroughness" requirement. When Sam and Sal make their on-site investigations of

Fred's ant colony, they are to be thorough. In this simple case, "thoroughness" (by stipulation) includes these two elements: (i) the investigation must be such as to carefully discriminate Ant A from Ant B (we would be idiots to leave that out); and (ii) the investigation must involve "sampling" the ant colony frequently enough and over a long enough time to stand a good chance of catching weak stochastic correlations between the properties of Ant A and of Ant B (since even if Ant A and Ant B are both F at a given moment, one does not expect this coincidence to be repeated frequently on other samplings, if the two ants are weakly correlated for F-ness). The sampling might rely partly on ant-colony laws or inductive generalizations, but presumably would include a good measure of just checking a lot. If Fred's confusion were between two types of physical quantity, both measurable only instrumentally, or between two types of civil liability, not "measurable" at all, these details of what counts as a "thorough" investigation would have to be replaced by details appropriate to the subject matter involved. If Fred's confusion were between two "objects" we do not even believe in (as with the story of the Successors and Predecessors), we would need to deviate even further from the paradigm we are setting forth here. But back to ants.

We issue some instructions to Sam and Sal. (In a real life case of confusion, "we" could not do such a thing.) What our instructions to Sam and Sal represent is a set of constraints on what items of knowledge (or informed belief) count as determining the semantic value of a given sentence and what items of knowledge or informed belief do not count. It is easier to discuss the constraints as "instructions" binding our imaginary authorities. Our first instruction is directed to Sam and Sal separately. Let's phrase it as an instruction for Sam.

Sam understands that he is providing Fred with advice about what beliefs to act on. Ordinarily, an adviser should make an effort to communicate to the advisee just how firm or shaky a given piece of advice is. The natural way to do that is to attach a probability to each sentence, perhaps in very informal language or perhaps in full dress. But we do not allow Sam to frame his advice that way. On what we will call "the first go-round," Sam is to mark a sentence "Yes" when he would pronounce it true, mark it "No" when he would pronounce it false, and mark it "Can't say" when he can't figure out whether it is true or false. Sam may not like to make crisp true/false/can't say judgments on sentences, preferring in his own thinking, and in his many learned articles on ant colonies, to give only refined statistical assessments. So when he follows our instructions,

he will be holding his own feet to the fire, demanding of himself that he traffic in truth-values rather than likelihoods. We give him no choice. The intuitive line of thought we are developing invoked the concept of "truth from a perspective," not the concept of "probability from a perspective"; we are trying to fix up that line of thought and make it presentable. The alternative of invoking probability from a perspective is well worth pursuing. We won't pursue it.

Our next instruction will not make sense until later. It is not so much an instruction as a job description. We only ask Sam and Sal to give advice on "atomic" sentences of Fred's ant-colony language, where an atomic sentence is one free of connectives, including quantifiers. For example, we ask Sam and Sal to make a judgment about "Charley is asleep at three o'clock," but not on "Charley is asleep at three o'clock *and* Charley is a bit under the weather today," nor on "*Some* ants are asleep at three o'clock."

Next we give an instruction to Sam and Sal *as a pair*. "They," collectively, bear the obligation this instruction imposes. The instruction has several parts. At issue is what should be done on the "second round" of recommendations concerning a given sentence.

(i) When Sam (Sal) decides on a "Yes" recommendation on the first round, and Sal (Sam) also decides on a "Yes" recommendation on the first round, the pair of them are to issue a joint communiqué for the benefit of Fred and any others who may have dealings with Fred's ant colony and who speak Fred's ant-colony language. That joint communiqué is to say: **Y.** We stipulate that **Y** has the meaning: authoritative opinion is *undividedly* in favor of acting on a belief in the sentence. Rule (i) obliges Sam and Sal to join forces in a recommendation to that effect.

(ii) When Sam (Sal) decides on a "Yes" recommendation on the first round, and Sal (Sam) decides on a "Can't say" on the first round, the pair of them are to issue a joint communiqué for the benefit of Fred and any others who may have dealings with Fred's ant colony and who speak Fred's ant-colony language. That joint communiqué is to say **Y.**

(iii) When Sam (Sal) decides on a "No" recommendation on the first round, and Sal (Sam) decides on a "No" recommendation on the first round, the pair of them are to issue a joint communiqué saying **N.** We stipulate that N has the meaning: authoritative opinion is *undividedly* opposed to acting on a belief in the sentence.

(iv) When Sam (Sal) decides on a "No" recommendation on the first round, and Sal (Sam) decides on a "Can't say" on the first round, the pair of them are to issue a joint communiqué saying **N.**

(v) When Sam (Sal) decides on a "Can't say" on the first round, and Sal (Sam) also decides on a "Can't say" on the first round, the pair of them are to issue a joint communiqué saying **?**. We stipulate that **?** means: authoritative opinion is silent as to whether one should act on a belief in the sentence.

(vi) When Sam (Sal) decides on a "Yes" recommendation on the first round, and Sal (Sam) decides on a "No" recommendation on the first round, the pair of them are to issue a joint communiqué saying **Y&N.** We stipulate that **Y&N** has the meaning: authoritative opinion is both in favor of and opposed to acting on a belief in the sentence.

Let's call these several parts of our instruction "Rule (i), . . . , Rule (vi)." The properties **Y, N, ?,** and **Y&N** ascribed to sentences by the joint communiqués are going to serve as our semantic values. The reader may have recognized these as the semantic values of Nuel Belnap's "useful four-valued logic."[4] All of Belnap's logical theory applies to our semantics, since that logical theory happens "below the level of" this particular interpretation of the four semantic values. The logical theory will provide an essential premise in the philosophical argument of this chapter, as will be clear soon enough. I will replicate almost none of Belnap's theoretical argument, referring the reader to his article, and other relevant sources. It is fortunate that this logical theory is on the shelf waiting, since an epistemologist who assigns herself the job of making a logic has a fool for a client.

Because **Y, N, ?,** and **Y&N** are epistemic properties of sentences, there is such a thing as a sentence being **Y&N,** and the value **Y&N** is sharply conceptually distinct from the value **?.** There is a difference between the authorities providing lots of information, but conflicting information, and the authorities providing no information. This point is Belnap's, and is at the heart of his own epistemic interpretation of the four values, as it is at the heart of mine.

Sam and Sal cannot investigate and comment on every atomic sentence in Fred's ant colony language; there are too many. There are two ways to go. We can let the value **?** do double duty, serving as the message "We can't tell even after investigation" in a joint communiqué, and serving as an indication that Sam and Sal haven't got around to investigating the sentence. Or we can let **?** do this double duty, but also try to extend the class of sentences Sam and Sal *effectively* "comment on" by assigning values whenever possible on the basis of a counterfactual judgment of what they would say if asked to investigate. If Sam and Sal were real people instead of

pretend people, so that we could know rather a lot of detail about how they investigate and what answers they come up with, we could evaluate a fair number of sentences by this counterfactual method. Any epistemic semantics faces some version of this problem, though it can get swept under the rug by talking as though it were a fact of nature whether a sentence is "highly confirmed," or "manifestly reasonable," rather than a fact about the actual epistemic assessments made by some (perhaps idealized) inquirer.

Now let's see which of these semantic values are indicators of "profitability for Fred" and which are indicators of "costliness for Fred." To keep our bearings, let's proceed by working through the four values by checking off one at a time Rules (i) through (vi), since taken together they specify how the individual opinions of the authorities Sam and Sal generate one or another of the four values in a joint communiqué. For simplicity let's concentrate on sentences of the form "Charley is F."

Consider Rule (i):

First suppose F-ness is a property like "sleeping (now)," discernible by any number of means, some of them noninferential, given a bit of care. When Sam and Sal both mark the sentence "Charley is sleeping" with "Yes," it is preposterous to suggest that neither big ant is sleeping. Sam and Sal would not be such renowned ant-colony authorities if they were that wrong about ants very often. Nor is it reasonable to think there is a "split," one of the big ants being asleep while the other isn't. If there were a split, Sam and Sal would have picked up on it, since each of them made a thorough investigation. (We ignore the vexed question how often ants wake up.) The upshot is that for such "easily accessed" properties as "asleep" a pair of "Yes" votes from Sam and Sal, encoded in their joint communiqué by the value **Y,** is a good reason to believe that both Ant A and Ant B will cooperate with Fred if he acts on the sentence so marked.

Not all properties are as epistemically accessible as "sleeping (now)" is. Some must be ascribed on the basis of inference from a combination of observation and theory. Some are ascribable on the basis of observation, but only after repeated observations. An example of the latter sort of property, well known to ant-colony authorities, is "being sniffle-prone." Sam and Sal might very well miss a divergence in sniffle-proneness between Ant A and Ant B on a one-shot examination, for instance, if both ants were sniffling disgracefully when observed. But the "thorough" examination Sam and Sal are required to make has a diachronic intensive-sampling aspect. If Ant A is not sniffle-prone while Ant B is, the joint communiqué

eventually issued on the sentence "Charley is sniffle-prone" should assign it the value **Y&N,** not the value **Y.** We must trust Sam and Sal to wait until they have done a fair amount of sampling before issuing communiqués on such sentences. Positing that they do wait, a pair of "Yes" verdicts from them, summarized in the joint communiqué as **Y,** is good reason for thinking that both Ant A and Ant B will cooperate with Fred if he acts on the sentence "Charley is sniffle-prone."

There are other forms of atomic sentences, and other types of predicate. The plausibility arguments we just gave are illustrative of the breed. Rule (i) requires that a pair of "Yes" votes from Sam and Sal on a given sentence be encoded by a **Y** in their joint communiqué. When the **Y** value arises in that way, one should expect *both* Ant A *and* Ant B to cooperate with Fred if he acts on the sentence. So far, the value **Y** shapes up as an indicator of profitability.

Consider Rule (ii):

Suppose Sam (Sal) marks a sentence "Yes" on the first round, but Sal (Sam) can't say. Then the only authoritative judgment is favorable. That is good reason to think that at least one of the big ants will cooperate with Fred if he acts on the sentence, and no reason to think the other will not cooperate. So when a **Y** value is assigned because of Rule (ii), the **Y** value is an indicator of profitability.

Consider Rule (iii):

A pair of "No" votes from Sam and Sal on the first round is good reason to think neither Ant A nor Ant B will cooperate with Fred if he acts on the sentence. The arguments we gave in connection with Rule (i) can be adapted for this case. So when the value **N** has been assigned via Rule (iii), it is an indicator of costliness.

Consider Rule (iv):

When either Sam or Sal marks the sentence "No" on the first round, but the other marks it "Can't say," the only authority to have spoken has come out against the sentence. So an **N** value assigned because of Rule (iv) also is an indicator of costliness.

Consider Rule (v):

Neither authority can say whether the sentence is true or is false (as each interprets it). So a **?** value is neither an indicator of profitability nor an indicator of costliness. Fred is on his own.

Consider Rule (vi):

When Sam and Sal divide their vote "Yes" and "No" on the sentence "Charley is F," it is reasonable to think that this has happened because Ant

A and Ant B diverge with respect to F-ness. What will happen if Fred relies upon such a sentence? He will meet with mixed success, and it is a reasonably sure thing that he will. This is not to be confused with the very different claim that there is no telling whether he will meet with success or failure, indicated by two "Can't say" votes, and summarized in the value **?.** We need to explain the meaning of "mixed success." Let's do it via an illustration.

Fred, always dabbling in ant-colony science, "tests Charley" to see if he likes radishes. Charley gobbles up the test radish. Fred leaves some radish scattered around the ant colony for a treat. He finds about a big ant's worth of it eaten. And he finds a horrible mess near each piece of radish, as though a big ant had been nauseated by the very sight of radish. Sam and Sal have done their research, thoroughly and professionally. Sam has marked the sentence "Charley likes radishes" with a "Yes," since Ant A would kill for a radish. Sal has written "No," having stuck radish beneath Ant B's snoot in diverse circumstances, always with horrific results.

Fred formulates some of his goals in terms of "Charley." For instance, he says "I hope Charley eats the radish." Since the sentence "Charley will eat the radish" is incapable of being either true or false, Fred is incapable of achieving the goal "Getting Charley to eat the radish," because he cannot make-true the sentence "Charley is eating (or ". . . ate . . .") the radish." But, as we know and Fred does not know, Fred is capable of causing Ant A to eat a fine meal of radish, just as he is capable of causing Ant B to feel ill at the sight of some radish. In fact he does both of these things.

Fred could evaluate these results (expressed in his terms) in either of two spirits. He could be guarded, reluctant to take a stand on what has happened. He might speak of how things "appear," or of what "may be" the case. For instance, he might say "Charley appears to have eaten the radish, but Charley also appears to have been made ill by the sight of it. I do not know what really happened." Alternatively, Fred might embrace all of the outcomes as real, and not mere appearance. He might say "Charley ate the radish, but he was made ill by it. I do not understand why an ant would—or even could—do that." If Fred opts for the second characterization, he is treating the events as constituting a challenge to his extant "ant theory." Neither of these alternative characterizations is the "right" one for him to adopt. The second probably would be the more fruitful, since it would be more likely to lead him to examine his concepts, including the concept "Charley." If he opted for the tentative "mere appearance" characterization, he would be less likely to make drastic conceptual revisions,

for there are many explanations for unexpected appearances, only one of which is that one has a fundamentally flawed theory.

What matters to us is that Fred would have the second option. He could embrace as real, and no mere appearance, the "paradoxical" outcome of his attempt to "feed Charley some radish." The question is how he should feel about that outcome. And the answer is that once again he has options. For example, he could take the "too astonished to know how to feel" option. That is, he could take the view that he did not envisage anything so bizarre when he set about scattering the radish. He expected one or the other of two coherent packages of events: radish eaten with no sign of distress (what he thought likeliest), or no radish eaten plus signs of at least "mild annoyance on Charley's part" (at having all that inedible stuff scattered around his dirt pile). The outcome of radish eaten plus signs of distress was not on Fred's list, so it has no place in his preference ranking.

But people often do not correctly identify the spectrum of possible outcomes of an action in advance; they must decide retrospectively whether the actual outcome—one they simply never anticipated—should count as "success," that is, should have a place in their reconstructed preference ranking. Fred may take this to heart and try to decide whether he is, or should be, pleased or displeased. Any number of considerations might weigh with him, but whatever they are, the "raw material" is there for him to come down either way. As Fred would put it, Charley has eaten the radish, a hearty big ant's worth of it. That is a desirable outcome. But Charley has been sickened by the radish, too. That is undesirable. One might outweigh the other, or they might balance. The important point for our purposes is precisely that the raw material is there, as Fred sees it, to add things up either way according to his applicable values. Let us call an outcome like this one "mixed success." An actor might describe a particular case of mixed success as a success (all in), or as a failure (all in), or as a wash. But a mixed success always contains some good as measured by the actor's (eventual) preference ranking—maybe overridden by the bad—as well as some bad—maybe overridden by the good.

We can predict with some confidence that *if* Fred adopts the attitude that the bizarre and paradoxical outcomes all are real and not mere appearance, and *if* Fred adopts the retrospective, after-the-facts-are-in, point of view when deciding whether "what he meant to achieve" has indeed been achieved, then Fred will find that the raw material is there to judge that he has done rather well, or to judge that he has done rather poorly, or to judge that it is a wash. *It would suffice* for making this prediction (for *us* to

make it) that Sam and Sal, respectively, mark the sentence "Charley likes radish" with "Yes" and with "No." They are authorities, and each has made a thorough investigation. Something must account for the divergence between them, and the likeliest account is that Ant A does, for sure, like radish, while Ant B for sure doesn't—or the other way around.

Putting all this together, we have:

(Y) The value **Y** is an indicator of profitability.
(N) The value **N** is an indicator of costliness.
(?) The value **?** is no indicator either way.
(Y&N) The value **Y&N** is an indicator of profitability and an indicator of costliness.

The plausibility arguments for (**Y**), (**N**), (**?**), and (**Y&N**) employed only atomic sentences as working examples. It will be good to keep that in mind when we get around to extending the reach of the four semantic values to logically compound sentences. If we were to extend the reach of **Y, N, ?,** and **Y&N** to compound sentences in such a way that it could no longer be held that the four values encode *authoritative* opinion, opinion "rationally derived according to standards appropriate to the subject-matter," there would cease to be any reason to think **Y,** for example, is an indicator of profitability.

Finally, we can draw a moral: our line of thought has been that someone who attributes confusion to someone else (or maybe to herself at another time) is "taking a semantic position" concerning the proper way to evaluate the reasoning of the person said to be confused.[5] This semantic position holds that the confused person's arguments ought to be appraised for validity by means of a "paternalistic" semantics. At the moment we are spelling out the structural details of a type of semantics, using one simple imaginary example for illustration. We have got as far as describing the basic semantic values, and describing how they are to be interpreted. When we finish with the details, our claim will be that in any given instance of confusion attribution, the semantics being recommended by the confusion-attributer is of this type.

But in order for a particular instance of this type of semantics to be described, reference must be made to particular real people, or perhaps communities of people, to play the roles of the semantic-value-determiners, the roles played by Sam and Sal. It must be clear what these people are like intellectually: in what ways they are reasonable and in what ways not; what their investigative methods are; in brief, what basis there is for calling them

authorities. This differs from subject area to subject area. In our imaginary examples, we can stipulate to the possibility of "ant-colony authorities." But in reality it will never do to stipulate, because there are many subject areas for which there cannot be any authorities, in a sense of "authority" that one can use to assess the profitability and costliness of beliefs. Whom could you consult to find out whether to act on a belief in astrology? Nothing will do but to have some exemplars of people who do possess the particular brand of authority, of trustworthiness, required to determine the semantic values. Sometimes a well-controlled extrapolation is acceptable: "mathematicians fifty years from now." But such cases are rare.

This is at once a limitation on the scope of this type of epistemic semantics—where the epistemic properties invoked must have implications for practice—and an epistemological security blanket. If no suitable real authorities can be pointed out or constituted from known real people, a semantics of this type cannot be given at all, and the attribution of confusion is contentless. This means that if you sit around wondering whether "all of your ideas might be confused," as did some of the early modern philosophers after Descartes scared them, you can take comfort, complete comfort, in the fact that since you really mean this to be a worry about the human condition, and since you cannot name your Sams and Sals, what you are wondering is ill-defined.

13

How Fred Should Think

When a sentence is **Y** or **Y&N,** let us say it is "at-least-**Y.**" When a sentence is **N** or **Y&N,** let us say it is "at-least-**N.**" Belnap's four-valued definition of a valid argument is:

(Valid) An argument is valid (a) if every premise is at-least-**Y,** then the conclusion also is at-least-**Y;** and (b) if the conclusion is at-least-**N,** then so is some premise.

Valid arguments preserve at-least-**Y,** and also preserve the absence of at-least-**N.** This is as it should be. We are explicating the concept "follows-from" in terms of practical profitability and costliness: we want Fred to reason in ways that preserve the success-promoting properties of his premises, when he is fortunate enough to have accepted such premises, and we want Fred to shun patterns of reasoning that might inject the property of being failure-promoting into his system of beliefs when it was not there before. The only measures of these things we allow are the authoritative judgments of Sam and Sal, so when Fred has a batch of **Y** premises, he has the good luck to have the weight of every authority on his side. We demand that he not fritter this prize away merely by his choice of methods of inference. When Fred has no **N** premises, we demand that he not infer an **N** conclusion, thereby recklessly flying in the face of all authority merely by his choice of a pattern of inference.

But the authoritative judgments made by Sam and Sal may differ, resulting in one or more of Fred's sentences having the value **Y&N,** offering him the promise of "mixed success" were he to go and act on the sentence. We do not want Fred to fritter away even this much **Y**-hood merely by reasoning ineptly; but we also do not want him introducing even this much **N**-hood, when it was not there already, merely by reasoning ineptly. So (for example) we do not allow Fred to pass from a batch of **Y&N** pre-

mises to a conclusion which is flat **N,** nor (for example) do we allow Fred to pass from a batch of ? premises to a **Y&N** conclusion.[1]

This justification of (Valid) depends on the fact that **Y** is an indicator of (and only of) profitability for Fred, while **N** is an indicator of (and only of) costliness. Fred should seek out **Y** sentences to believe regardless of how he evaluates situations of mixed success, and he should avoid believing **N** sentences, again regardless of how he evaluates situations of mixed success. When a sentence is **Y&N,** Fred could rationally go either way—worse off than when believing a **Y** sentence, though not as badly off as when he believes an **N** sentence. If we find ourselves confronted with a case where the connections between the value **Y** and profitability and between the value **N** and costliness cannot be shown to exist, this justification of (Valid) cannot be given. If we made an attribution of confusion in such a case, our attribution would misfire: we would be recommending that some person's reasoning be appraised by means of a criterion that we could not show to be a *validity* criterion.

We must be careful how we understand such words as "Fred should seek out **Y** sentences to believe." Assuming Fred is not especially knowledgeable about the science of logic, he cannot be expected to "seek out **Y** sentences to believe," or to "see that he employs arguments which preserve being-at-least-**Y,**" if this is understood as meaning that he selects arguments according, in part, to whether they are being-at-least-**Y**-preserving. Similarly, he cannot be expected to select arguments because they are not-being-at-least-**N**-preserving. But—to switch semantic paradigms—most people do not know how to select arguments that preserve truth, in the sense of knowing how to select arguments because they possess that semantic property, and most logical thinkers have never stopped to notice that they like arguments that preserve truth. What we expect people to do is, first, select arguments which as a matter of fact do preserve truth, and second, be moved by the fact that an argument does or does not preserve truth in the event that such a fact is brought to the person's attention and clearly explained. We expect people to preserve truth, and *care,* in principle, whether they preserve it.[2]

That is how we should understand the expectation that Fred obey (Valid). We expect him to select arguments which as a matter of fact do preserve being-at-least-**Y,** and do preserve not-being-at-least-**N,** and we expect him to be moved by the fact that an argument has (or lacks) that semantic property in the event it is clearly explained to him that the argument does have it (or lacks it). Now, the semantic values **Y, N, Y&N,** and

? are defined in terms of the advice proffered by Sam and Sal, each of whom can distinguish Ant A and Ant B, and each of whom uses the name "Charley" unconfusedly. Fred has not got any notion that such knowledge and such linguistic practices are possible. Therefore Fred cannot understand the grounds upon which Sam and Sal make their expert judgments and indeed cannot even understand the language in which they do it. Fred cannot appreciate the "ultimate theoretical origins" of assignments to sentences of the semantic properties being-at-least-**Y** and not-being-at-least-**N.** Nevertheless, the condition (Valid) is *authoritative* for Fred.

The stipulation that Sam and Sal are experts concerning Fred's ant colony guarantees that Fred should want to reason in ways that preserve the good advice these experts toss his way. It does not matter that Sam and Sal know things Fred does not know, such as that there are two big ants. We expect experts to have a different take on things than we have, often a radically different, more sophisticated take. Experts may well reject the descriptive language we use in favor of a different language we cannot understand at all, and which, we rightly suspect, enables them to draw distinctions we do not draw and maybe cannot conceive how to draw. This does not alter the fact that we should act on the advice of these experts to the extent possible. With Fred, Sam, and Sal, we have a very simple case, a toy case, of this system of epistemic relations.

One of the virtues of an epistemic semantics of the type we are considering is that it connects its subject (the person whose reasoning is at issue) with the world *via the thought* of certain designated experts. These experts may or may not themselves be making good contact with reality, but they are making better contact than our subject is making. They are the instruments our subject uses to amplify and clear the noise from nature's signals. An epistemic semantics makes essential use of the idea of justified trust in one's instrumentation. That trust is justified even when you do not know how the thing is made, provided you know the brand. Fred does not know exactly how Sam and Sal "are made," but his trust is justified anyway. The metaphor is not perfect, of course; Sam and Sal do not come with a brand label attached, so Fred's trust must be justified on other grounds. We have spent some time looking at those grounds.

The final communiqués issued jointly by Sam and Sal assign one or another of the values **Y, N, ?,** or **Y&N** to atomic sentences. Let's decide how the four values are to be assigned to logically compound sentences. We are going to take that job out of Sam's and Sal's hands. Our strategy will be to assign semantic values to logical compounds by seeing what assignments it

would be *logically cogent* to make. Later we will be obliged to back off this strategy just a little, so let's get clear why it is a good strategy. Then we'll know what price we pay for deviating from it.

We have stipulated that the opinion of an authority be rationally derived by standards appropriate to the subject-matter. Therefore Sam and Sal—whoever they are for a given case of confusion—have seen to it that the semantic values assigned to atomic sentences reflect opinions that meet this requirement. (We may or may not have used a "counterfactual" method for assigning values other than **?** to atomic sentences Sam and Sal never thought about, but if we have done so, those values still are supposed to be "the ones Sam and Sal would assign.") It is up to us to see that we don't fritter away the "rationality" of Sam and Sal's opinions when we extend the range of the semantic values to logically compound sentences. We surely will fritter it away if we ignore logical cogency. Happily, there is no other way for us to fritter away the rationality of Sam and Sal's judgments, since we need apply no subject-matter-specific knowledge of our own. The stakes are high here: if we were to lose the right to think of the semantic values as rationally ascribed, we also would lose the right to think of the values as indicators of profitability or costliness (or mixed success in the case of **Y&N,** or nothing at all in the case of **?**). That would take away our defense of (Valid).

Notice that there is no circularity in the idea that we demand logically cogent value assignments to compound sentences. There is no assumption that we already know how to "apply logic" to a class of sentences, as a precondition for stating what their logical structure is. The semantics we are giving is a "logic" for Fred's ant-colony language, whereas the logical-cogency judgments we must make to decide how semantic values should be assigned are logical judgments about semantic metalanguage sentences, such as the sentence "When A and B both have the value **Y,** the conjunction of them also (logically) should have the value **Y.**" The semantics we go on to construct, on the other hand, applies to sentences like "Charley smells bad and Charley sleeps a lot."

What we want to do, in effect, is decide what "communiqués" it *would be* logically cogent for Sam and Sal to issue jointly. It continues to be their *collective* advice we care about, since this is our way of getting a semantics that does not pay attention to which big ant does what. Even when a sentence has the value **Y&N,** that value does not encode any information about which ant accounts for which vote.

We all have considerable experience at making evaluations of joint re-

ports, including evaluations of their logical cogency. One new wrinkle shows up when one considers a joint report, such as a position paper issued by a team of researchers: the value **Y&N.** This wrinkle *always* shows up, it is not just pulled out the air for purposes of constructing this semantics. There is nothing unfamiliar about a group of authorities issuing a "some of us say yes, some of us say no" opinion on some matter. Of course a particular group of authorities might have special rules, like our Rules (i) through (vi) of Chapter 12, prescribing how individual "yes" and "no" opinions are to be combined into a joint statement. Those rules might specify that a small minority of "no" opinions are to be disregarded for the purpose of making a collective statement. Or they might let a single "no" opinion amid hundreds of "yes" opinions generate a "some say yes, some say no" collective statement (that is what we have told Sam and Sal to do). But once these rules have been established, once we have got past the issue of how to combine individual judgments into collective judgments, the logic of "some say yes, some say no" collective judgments is (mostly) straightforward. For example, if Sam and Sal (jointly) say **Y** to sentence A, and (jointly) say **Y&N** to sentence B, they better (jointly) say **Y&N** to the conjunction "A and B."

Just as "some of us say yes and some of us say no" is a familiar option for groups to take when making collective pronouncements, with a familiar logic at which all of us are practiced (or, if you prefer, about which all of us have logical intuitions), the response "don't know" is familiar when opinions are solicited. Since ours is an epistemic semantics, with semantic values representing opinions, there is a "don't know" value. Nobody has any trouble making judgments of logical cogency about mixtures of "yes," "no," and "don't know" statements, either. Here it is irrelevant whether we are dealing with individual opinions or collective opinions. Example: if Sam and Sal give a **Y** to sentence A and a **?** to sentence B, they better give a **?** to the conjunction.

We make no attempt to provide an external justification for our cogency intuitions about how the **?** typical of epistemic reports, and the **Y&N** typical of collective reports, mix with the other (epistemic and collective-report) values **Y** and **N.** In particular, we will not try to show that this pattern of intuitions can be derived in some fashion from facts about how the metaphysical (and hence nonepistemic) properties truth and falsehood mix in a two-valued semantics. To require of ourselves that we do such a thing would be like requiring of ourselves that we derive our ordinary moral principles from the code of conduct the Geneva Conventions im-

pose upon prisoners of war. The validity standard of classical logic can be defended as a good choice for certain purposes, just as the validity standard we are considering here can be defended as the best choice for certain purposes. Neither of these validity standards captures "what following-from really is."

Let's confine ourselves to worrying about the three connectives "not," "and," and "or." Fred has many other devices for introducing logical complexity, but we'll ignore them. That limits the scope of any philosophical claims made on the basis of the semantics. But it will be enough to illustrate the *type* of philosophical claim one can make.[3] And it will give us a picture of valid inference just rich enough to permit some useful philosophical application in later chapters.

Let's follow to the letter Belnap's recursive rules for assigning semantic values to logical compounds. There are three rules, one for each of the three connectives. The rules can be expressed in the form of "value tables." Here is Belnap's value table for negation:

A	?	N	Y	Y&N
not-A	?	Y	N	Y&N

And here is the table for conjunction (possible values of conjunct A vertically at left, possible values of conjunct B horizontally at top, resulting values of the conjunction "A and B" in the 4×4 grid):

	?	N	Y	Y&N
?	?	N	?	N
N	N	N	N	N
Y	?	N	Y	Y&N
Y&N	N	N	Y&N	Y&N

And, finally, the table for disjunction (possible values of disjunct A vertically at left, possible values of disjunct B horizontally at top, resulting values of the disjunction "A or B" in the 4×4 grid):

	?	N	Y	Y&N
?	?	?	Y	Y
N	?	N	Y	Y&N
Y	Y	Y	Y	Y
Y&N	Y	Y&N	Y	Y&N

Let's start with the recursive rule for negation. If the sentence A has the value **?**, this means that the authorities, both of them, can't say, perhaps

because it is too hard a problem for them, perhaps because they haven't worked on it enough yet. Either way, the report on not-A, were one to be issued, better read "Can't say."

If the sentence A has the value **Y,** then its negation better get an **N.** And vice versa. These two results are *not* due to the values **Y** and **N** being "truth and falsehood under another name." They are not truth and falsehood under any name. The results are due to an obvious necessary condition for having a rational practice of affirmation and denial: one may not affirm the negation of what one affirms, and if one chooses to affirm or deny that negation, one must deny it (and the other way round for the negation of what one denies). This is not the same thing as saying: one may not ascribe truth to the negation of what one has ascribed truth to, and if one chooses to ascribe either truth or falsehood to that negation, one must ascribe falsehood (and the other way round for the negation of what one has ascribed falsehood to). There may be a connection between this last principle, a constraint on ascribing the properties truth and falsehood, and the former principle, a constraint on affirming and denying. But they are not the same principle. I suspect they sometimes strike philosophers as the same principle, because if one were using the words "true" and "false" prosententially, they would be the same principle—for in that case the words "true" and "false" would be devices for affirming and denying, nothing more. But in semantic theory one uses "true" and "false" to express the properties truth-of-the-world and falsehood-of-the-world; one does not use them prosententially.

If the joint report on A is **Y&N,** the report on not-A better not represent Sam and Sal as jointly saying plain yes, or saying plain no, or punting with a "can't say." They, collectively, must say yes and no again. (Imagine a team of scientists saying in a joint research report: "We cannot reach unanimous agreement on the matter of health risks from very-high-voltage longlines. Some of us believe there is a risk, some of us deny there is a risk. However, we all agree that there is no risk." One would suspect a typo.)

That justifies the rule for negation.

Conjunction and disjunction oblige us to survey more combinations, but until we hit the place where we must abandon intuition for logical theory, they are just as simple, combination by combination. Let's have a look at conjunction. We'll skip the details for disjunction, since the general routine is the same: smooth intuitive sailing except for occasional spots where we allow theory to prevail. Belnap has two simple rules which, taken together, completely determine the values in the table for conjunction:

(I) Mark "A and B" with at least one **Y** just in case both A and B have been marked with at least one **Y.**
(II) Mark "A and B" with at least one **N** just in case at least one of A and B have been marked with at least one **N.**[4]

The elegant way to proceed would be to argue directly for these two rules; the inelegant way would be to work through the table one entry at a time. Let's mix the two, using the two rules as a roadmap of the table's structure, counting our job done when both rules are justified.

(1) Suppose A is **Y** and B is **Y.** Logical cogency requires that Sam and Sal jointly give the value **Y** to the conjunction "A and B."
(2) Suppose one of A or B is **Y** but the other is **Y&N.** Logical cogency requires that the conjunction be given a **Y&N.**
(3) Suppose both conjuncts get a **Y&N.** Logical cogency requires that the conjunction get a **Y&N.**

This case is worth worrying about. There is no problem here when the **Y&N** values for A and for B arise because Sam says "Yes" to both A and B while Sal says "No" to both, or when Sam says "No" to both while Sal says "Yes." But suppose Sam says "Yes" to A and "No" to B, while Sal says "Yes" to B and "No" to A. Sam and Sal would each say "No" to the conjunction "A and B." But this is a case where Sam and Sal would issue the collective report **Y&N** for A and would issue the collective report **Y&N** for B. Does logical cogency require that we assign the value **Y&N** to the conjunction "A and B" even in this case, where both Sam and Sal taken separately would say "No" to that conjunction? Yes, it does—sort of. One of our goals is to describe a concept of validity for Fred's reasoning which is "blind to the distinction between Ant A and Ant B." We will accomplish that, in part, by employing semantic values that do not encode any information about which big ant has which property, in the event that the ants differ in their properties. So we use the value **Y&N,** a semantic value that does not discriminate between Sam's (Ant-A-based) input and Sal's (Ant-B-based) input. We could switch to a five-valued logic, distinguishing between the semantic values [Sam: yes; Sal: no] and [Sam: no; Sal: yes]. But then our semantic values would encode at least some differential information about the two big ants. If we restrict our attention to semantic characterizations obeying this "nontransparency" requirement—you can't "see through" the value **Y&N** to find out which authority says what—then we must simply ask what to do with a pair of collective reports of the

form "sentence A: **Y&N**" and "sentence B: **Y&N,**" for which we are *ignoring the very possibility of* seeing through to the respective opinions of the two authorities. Surely it is most cogent to classify the conjunction "A and B" as itself **Y&N.**

(4) Therefore, if A and B each has been given at least a **Y,** the conjunction should be given at least a **Y.**
(5) Suppose the conjunction "A and B" gets a **Y.** Logical cogency requires that each conjunct get a **Y.**
(6) Suppose the conjunction "A and B" gets a **Y&N.** Logical cogency would rule this out if either conjunct got an **N;** and likewise if either conjunct got a **?.** So the conjuncts must get either a **Y** or a **Y&N.**
(7) Therefore, if the conjunction "A and B" gets at least a **Y,** both conjuncts get at least a **Y.**

From (4) and (7) we have the rule:

(I) Mark "A and B" with at least one **Y** just in case both A and B have been marked with at least one **Y.**

(8) Now suppose one of A and B gets an **N.** Logical cogency requires that the conjunction get an **N.**
(9) Suppose one of A and B gets a **Y&N.** Four cases arise: (a) Both A and B get a **Y&N.** We settled this under (3) above; the conjunction gets a **Y&N.** (b) One of A and B gets a **Y.** We settled this under (2) above. The conjunction gets a **Y&N.** (c) One of A and B gets an **N.** This is case (8). The conjunction gets an **N.** (d) One of A and B gets a **?.** Logical cogency requires that the conjunction get a **?.** *We disregard logical cogency* and give the conjunction an **N.** Summing up the results of (9), if one of A and B gets a **Y&N,** the conjunction gets at least an **N.**
(10) From (8) and (9) together we have the result that if one of A and B gets at least an **N,** the conjunction gets at least an **N.**
(11) Now suppose the conjunction "A and B" gets an **N.** If both conjuncts were **?,** logical cogency would require the conjunction to have been given a **?,** not a flat **N.** If one conjunct were **Y** while the other was **?,** the conjunction should be **?.** And if both conjuncts were **Y,** the conjunction should be **Y.** So once we know the conjunction is **N,** we know one conjunct or the other should get at least one **N**—that is, either be **N** or be **Y&N.**
(12) Suppose the conjunction "A and B" gets a **Y&N.** That "No" vote

must have come from somewhere. So one of the conjuncts must be at least **N.** (In fact an exercise in checking off all the possibilities shows that one conjunct must be **Y&N.**)

(13) From (11) and (12) we have: If "A and B" has been given at least an **N,** then one of A and B has been given at least an **N.**

From (10) and (13) we have the rule:

(II) Mark "A and B" with at least one **N** just in case at least one of A and B have been marked with at least one **N.**

We'll skip giving a detailed defense of the value table for disjunction. Observe, though, that not all of the value table for disjunction is consistent with logical cogency. For example, when one disjunct has the value **?,** and the other the value **Y&N,** the disjunction has the puzzling value **Y.** It is time to explain the rationale for these deviations from the dictates of logical cogency, and likewise for the deviations in the table for conjunction (the conjunction of a **Y&N** sentence with a **?** sentence gets the value **N,** rather than the value **?**).

Let's take a step back from the details and recall our philosophical agenda. This was the background: some of Fred's friends who are privy to the facts about Ant A and Ant B ("we" in the story) choose to be judgmental of him rather than humoring. Our critical comment, "Fred thinks Ant A is Ant B," is a typical expression of this judgmental attitude. It is a semantic comment, not a psychological comment; the speaker is not ascribing a mental state to Fred, she is taking a semantic position on how Fred's reasoning should be appraised for validity. The speaker is plumping for inferential charity.

The comment that Fred thinks Ant A is Ant B contains almost no semantic detail. Still staying at an intuitive level, one can try various reformulations that are cast in explicitly semantic language—perhaps less likely to mislead than using "thinks" to make a semantic point. One can say that the idea is to allow Fred's arguments to count as "perfectly logical" even when they "fail to take account of the difference between Ant A and Ant B." This is suggestive, but no more than suggestive: after all, a person who knew nothing about the ant colony, and had no names, confused or unconfused, for any of the ants, presumably would "fail to take account of the difference between" Ant A and Ant B in her inferences. We, the philosophers, undertook to remedy this lack of semantic detail on behalf of ourselves, the confusion-attributers in the story.

We (reader and author, not we in the story) have had a go at describing in some detail a suitably charitable account of argument validity. (There may be others; so the meaning of the unphilosophical confusion-attribution that "Fred thinks Ant A is Ant B" should be thought of as underdetermined.) We were led to Sam and Sal, the authoritative sources of ant-colony opinion, whose joint communiqués serve as reasons why Fred ought to act on this ant-colony belief rather than some other ant-colony belief. Fred sometimes cannot appreciate how good these reasons are, but that is the way things usually are—people do not know why good advice is as good as it is. They must trust the adviser (or perhaps the adviser's credentials). Fred represents an extreme case: it is not much exaggeration to say that Fred is not conceptually equipped to appreciate how good Sam and Sal's advice is.

When Sam and Sal issue a joint report on some sentence, the results of two authoritative inquiries in which the name "Charley" is used to mean Ant A and Ant B, respectively, are merged. The semantic values **Y, N, Y&N,** and **?** encode this merged information. Since our Belnapian validity criterion is formulated in terms of these semantic values, our semantics captures part of what is meant by the intuitive idea that one should allow Fred's arguments to count as "perfectly logical" even when they "fail to take account of the difference between Ant A and Ant B"—it captures the "even when they fail to take account of the difference" part. But what about the "perfectly logical" part? What standard are we using to identify a "perfectly logical" argument when Fred gives one?

Imagine that we (in the story) are listening to Fred talk. We hear him give some arguments, all of them invoking "Charley." "Charley is strong and brave," he says, "and since he is brave (etc.)." A conjunction elimination has just gone past, and we smile. Fred's conclusion did indeed follow from his premise. He should be given high marks, from a logical point of view, for that tiny swatch of reasoning—so we say. It is clear, mostly, which of the arguments Fred gives, or might give, we would want to characterize as examples of logical reasoning. They are just the arguments we would characterize as "logical" if we were to give them, but substituting "Ant A" uniformly for "Charley." Or substituting "Ant B" uniformly for "Charley." Or substituting "Bill Clinton." Or "Epsilon Aurigae." The arguments (argument forms) we want to count as valid when Fred employs them are the usual suspects. We want to be able to say that as far as "being logical" goes, Fred is just like us; his confusion is not a result of illogicality. It's just that in order to give him credit, one must apply a validity criterion

one would not wish to apply to oneself. By recommending that it be applied to Fred we hold him to be semantically abnormal, semantically defective—in a word, confused. But we hold him to be smart.

It follows that the semantics we are describing must ratify, as "valid," those forms of argument we think of as characterizing completely logical thought in the normal case, in ourselves. If it does not do so, then however interesting it might be for one or another philosophical purpose, it will not suit the philosophical agenda here.

At the same time, it must be reasonable to claim that the semantics constitutes an explication of the concept "follows from." Suppose, contrary to fact, the recursive rules for assigning semantic values to logically compound sentences did so in a generally irrational fashion. Then there would be no basis for claiming that the semantic values attached to logically compound sentences reflect what (rationally) ought to be authoritative opinion concerning those sentences. There would be no reason in general to claim that being-at-least-**Y,** for example, *should be preserved by* a "truly logical" inference.

The world, as noted earlier, is not here to make life easy for philosophers. There is a conflict at the margins between the two desiderata:

(1) ratify as valid-according-to-the-semantics exactly the class of argument-forms we think are the "completely logical" ones, and ratify them even when they include "Charley" talk;
(2) make sure the theoretical concept "argument that is valid-according-to-the-semantics" is a good explication of the intuitive concept "argument such that the conclusion follows completely logically from the premises."

Given that there are alternative, equally acceptable explications of the intuitive concept "argument such that the conclusion follows completely logically from the premises," we will have satisfied desideratum (2) if we have specified *one* of these alternative explications.[5]

As we cannot fully satisfy both (1) and (2), we must make a choice. The choice I have made (it was Belnap's choice, though he did not put it in exactly the same terms) is to sacrifice (2) at a few points in order to nail (1); in order to get a semantics which ratifies as "valid" by its lights a class of inferences one can comfortably call "exactly the inferences we think are completely logical to make."

For instance, we let the value of a conjunction be **N** when it has a **?** conjunct and a **Y&N** conjunct, rather than letting the value of the conjunc-

tion be ⥽, as strict adherence to logical cogency when interpreting the connectives seems to require. This counterintuitive choice slightly weakens the fit between the concept "argument that valid-according-to-the-semantics" and the concept "argument such that the conclusion follows completely logically from the premises." But it buys us a tight fit between the class of arguments valid-according-to-the-semantics and the class of arguments intuition identifies as the completely logical ones for us to give, or for Fred to give. Belnap's logical theory, to which we may help ourselves, tells us what class of implications are valid according to our semantics *when we posit exactly the recursive rules we finally did posit*. The class of valid implications turns out to be the implications of the Anderson-Belnap Entailment system, specifically **E**fde.[6] It is reasonable to think the entailments belonging to this class are exactly the implications a completely logical person would be willing to accept (restricted to the sublanguage of Fred's language built from atomic sentences and the three sentence connectives). Roughly, the entailments of **E**fde include almost all of the classically (two-valued) valid implications, leaving out some implausible candidate entailments admitted by classical logic, as well as some candidate entailments which are rather more plausible. For instance, the silly implication:

A ⥽ not-A

is invalid in **E**fde, just as it is, and should be, in classical logic. The implication:

(A and not-A) ⥽ B

or "anything you wish follows from a formal contradiction" is a valid implication of classical logic, but is not a valid implication of **E**fde. It is hard not to see this as a theoretical strength of **E**fde as compared with classical logic; we would not like to see Fred, or anybody, saying "Clearly it follows from the proposition 'snow is both white and not white' that Bill Clinton is President." As I just said, there may be "rather more plausible candidate entailments," inferences we should regard as valid, and which are so regarded by classical logic, but which are not validated by **E**fde. More on that matter in a moment.

It is no surprise that there should be some trade-off between making an exact capture of the pretheoretical concept "follows-from," and getting the semantics to crank out as "valid" a class of inferences that fits our agenda.[7] Such trade-offs are typical of explications, although usually they

take a different form: the set of objects to which the explicated concept applies is pared down, or pared down in one way and augmented in another, as compared with what "intuition" dictates. In the case at hand that would mean letting the *class of* "valid-according-to-the-semantics" inferences differ from the class of inferences intuition teaches us are such that their conclusions follow from their premises. But that is just what our philosophical agenda will not allow. We need the class of inferences counted as valid by our explicated validity concept to be as nearly as possible the set of inferences that *strike us as* completely logical. So we find the play elsewhere: we let the fit between the metatheoretic *concept* "valid," and the intuitive logical *concept* "following-from" which it explicates, be less than perfect, though not in a way that significantly affects the match between the *extensions* of those concepts.

Still, we need to be clear about just how limited the reach of this semantics is. It is a semantics for inferences in negation, conjunction, and disjunction. Other sentence connectives can be defined in terms of these. But a feature the logic **E**fde shares with many other related logics makes for a problem the reader who is familiar mainly with classical logic may not notice. In **E**fde the argument-form "disjunctive syllogism" is not valid. The implication

not-A and (either A or B) ⇒ B

is not valid, though special cases of it are. This means that one cannot define a conditional, an "if . . . then," in the way one is taught to do in standard logic texts: by calling the schema

either not-A or B

the "material" conditional, rewriting it as

if A then B,

and then counting on the validity of disjunctive syllogism to guarantee the validity of "modus ponens," the absolutely fundamental principle of reasoning:

A and (if A then B) ⇒ B.

One gets to *modus ponens* from disjunctive syllogism by the steps:

(1) not-A and (either A or B) ⇒ B
[disjunctive syllogism];

(2) not-not-A and (either not-A or B) ⇒ B
[letting "not-A" be the sentence "A" in (1)];
(3) A and (either not-A or B) ⇒ B
[substituting A for the equivalent not-not-A in (2)];
(4) A and (if A then B) ⇒ B
[by the definition of "if . . . then" given above].

Even if this way of defining "if . . . then" has its drawbacks, and it does, at least we get *modus ponens* as a valid principle of reasoning. But, as **E**fde does not contain disjunctive syllogism, our little chain of justification for *modus ponens* is broken.

This means that if Fred gives the argument "Charley is annoyed; if Charley is annoyed he may bite my hand; so Charley may bite my hand," we cannot follow the classical approach and think of Fred's *modus ponens* as a disguised disjunctive syllogism, and therefore a valid argument. But our thesis is that when we (we in the story) attribute confusion to Fred, this is semantic position-taking; the whole point of this chapter's exercise in semantics construction was to fill in some details of a semantic position with which the confusion-attributer could comfortably identify. If the semantics we have constructed cannot even ratify a *modus ponens* when Fred gives one, surely one cannot claim with a straight face that we have shown how to represent Fred as being "just as logical as we are."

There is merit in this complaint, though somewhat less than might first appear. The semantics constructed in this chapter does not (as it stands) have the resources to ratify a *modus ponens* as valid. But that is because it does not have the resources to represent a *modus ponens* at all; not because it does have such resources but, disastrously, it classifies *modus ponens* as *invalid*. (Compare: it is easy to describe an "equivocation semantics" for Fred's reasoning, within which semantics (a) one can represent conjunction elimination, and (b) some conjunction eliminations are invalid—see Chapter 5.) It would be a good idea to elaborate our Belnapian semantics so that it applies to reasoning containing "if . . . then" sentences, *and* ratifies *modus ponens* as valid. I won't try.[8]

As things stand, we have a semantics that applies to a relatively small fraction of the arguments Fred may be expected to give. There is no way to be certain how smoothly it will generalize except to go ahead and work out the details of the generalization, though we can, and will, make some fairly plausible guesses. What our semantics gives us is a concrete example of a cluster of semantic concepts that seems adequate to represent the va-

lidity or invalidity of reasoning conducted in confused language, albeit a limited range of reasoning. Most of the philosophical applications discussed in the remaining chapters of this book rest, ultimately, on the character of these semantic concepts, and in particular on the ways they differ from Truth, Falsehood, and related classical semantic concepts.

I have no reason to believe that the only adequate semantics of confused reasoning is the one sketched here. Indeed, I believe several other approaches have promise and ought to be explored. This means that when we-in-the-story attribute confusion to Fred, our remark underdetermines the details of the semantic position we are taking. Time, and study by logicians, will tell. For the remainder of this book I will pretend the issue has been settled: when we-in-the-story attribute confusion to Fred, we are insisting that his inferences be evaluated for validity in terms of the epistemic semantic values **Y, N, Y&N,** and **?,** and by means of the criterion of validity described in this chapter (or some plausible elaboration)—although we-in-the-story, as nonspecialists, could not put it in those words. I must pretend *something*, because without some semantic details we will make no philosophical progress.[9]

VI

CURING CONFUSION

14

Semantic Self-Awareness

If there were a state of mind of "thinking Ant A is Ant B," the way to remove Fred's confusion would be to cause that state of mind to be replaced by another; perhaps by the state of mind of "realizing that Ant A is not Ant B." A good way to do that might be to point Fred toward Ant A, cavorting on the surface of the ant colony, and then, as Fred watched, haul Ant B up from the depths, placing it right beside Ant A. But since my thesis is that the confusion-attribution "Fred thinks Ant A is Ant B" does not ascribe any state of mind to Fred, I cannot tell quite so simple a story about how we might eliminate Fred's confusion. Even this form of words —"eliminate Fred's confusion"—is misleading, implying as it does that there is a person-state of some kind one should call "confusion." On my view of the matter, in order for it to become incorrect to say "Fred thinks Ant A is Ant B," and correct to say "Fred realizes that Ant A is not Ant B," the reasons we have for taking a certain semantic position must be undermined or overridden. It must come to pass that—in our view—an attitude of inferential charity toward Fred has become unreasonable, giving way to the contrasting attitude that if Fred makes an inference that fails to respect the difference between Ant A and Ant B, that inference of Fred's ought to be evaluated as invalid. A normative illness requires a normative cure.

Obviously, putting Ant A and Ant B side by side before Fred's eyes could lead to it being reasonable of us to say "Fred realizes that Ant A is not Ant B." So any account of the shift from one semantic position to the other must explain the role played by this, and other, empirical experience of two big ants Fred might have. But the role played by such empirical experience cannot be merely that it provides Fred with good grounds for altering his belief that there is a single big ant, or with good grounds for altering his belief that the ant he has observed doing such-and-such things is identical to the ant he has observed doing thus-and-so things. Some of the

arguments deployed in Chapter 3 to show that Fred's confusion is not a matter of his having such false beliefs can be altered slightly to show that he could be given evidence that those beliefs are false, and alter them in response to that evidence, without this making the slightest difference to the reasonableness of our confusion-attribution. For example, a person Fred has every reason to trust could play a trick on him by showing him a third big ant, claiming it lived in the ant colony in a small cave Fred had never noticed. Or do whatever it takes to get Fred to reject the belief that there is just one big ant in the ant colony, and reject it with good reason. We would continue to say Fred thinks Ant A is Ant B. Similar arguments can be constructed for other beliefs Fred has, to show that he can be given good grounds for rejecting these beliefs without our coming to have a basis for retracting our attribution of confusion. I won't belabor the point.

The proper account of how it comes to be reasonable of us to withdraw our attribution of confusion is simple, but there is no way to state it clearly until we have put some further epistemological machinery in place. Let's start.

When the real variable r is restricted to numbers with an absolute value less than one, the infinite power series

$$(\mathrm{PS})\ 1 + r + r^2 + r^3 + r^4 + \ldots$$

converges and is equal to $1/(1 - r)$. When $r = 1$, the ratio $1/(1 - r)$ is undefined. What happens in other cases, for instance when r is greater than one? For a value of r greater than 1, the series you get by plugging that value of r into (PS) diverges. But the result of dividing 1 by $(1 - r)$ for your chosen value of r is some real number or other. So it seems that the series $1 + r + r^2 + r^3 + r^4 + \ldots$ cannot be equal to $1/(1 - r)$ for $r > 1$. Standard theory ratifies this intuitive argument. Let us assume, then, that it is a mistake to identify the series $1 + r + r^2 + r^3 + r^4 + \ldots$ with the quotient $1/(1 - r)$ for $r > 1$.

But here is an argument to the effect that the series (PS) is equal to the quotient $1/(1 - r)$ for all r except 1, in which case the quotient is undefined: high school algebra tells us how to divide 1 by the quantity $(1 - r)$ whenever the division is meaningful; that is, in all cases except $r = 1$. As you carry out the division, you get term after term of the series (PS), for as long as you keep dividing. Wherever you stop, there will be a remainder term, but if you start dividing again you will get still more terms of the series. It is easy to confirm by a simple proof that if, "in principle," you kept dividing without stopping, the series (PS) would keep on flowing out of your computational mill, term by term. From these considerations we

draw the conclusion that in all cases except $r = 1$ the series $1 + r + r^2 + r^3 + r^4 + \ldots$ is equal to $1/(1 - r)$, since the series is the "result" of dividing 1 by $(1 - r)$.

In order for this argument to hit you the right way, you must begin by imagining a student who has recently learned how to perform divisions of the sort involved here, with this difference: the examples have been chosen so that the "answer" always is a (finite) polynomial. You also must imagine that the student never has heard of the idea of an infinite series converging or diverging. She is, however, a very good student, imaginative and quick to pick up on new applications of familiar methods.

Do the rules of the high school algebra game actually specify how to tell the result "in principle" of an "infinitely repeated division"? Let us assume the answer is no. There is no good reason for the "synthetic division" algorithm to be accompanied by such logical subtleties at that level of mathematical education, and plenty of reason for it not to be. If one wrote out the inference rules of the high school algebra game in as much detail as possible, using a predicate "follows-from" to encode the licensed inferences, the predicate would be vague. The rules for using the predicate would not determine whether the conclusion of our argument follows from its premise, or does not follow. At some points in its history, science has resembled the high school algebra game in this respect. But that is getting ahead of the story.

When the skilled student player of the high school algebra game confronts our argument for identifying the series (PS) with the quotient $1/(1 - r)$ for all r except 1, the rules of the game do not give her direct and explicit guidance. But they do give her *some* guidance. Think of it this way: she must decide upon an analogical extension of the rules to cover the new case before her, an analogical extension into the vagueness zone of "follows-from." Our little argument rests upon an intuitively plausible analogical extension, so the argument has force. Many ordinary inferences are like this one; an inferential practice is extended, relying on analogy with "standard" or "core" examples of the practice, in order to fit the practice to somewhat novel conditions, in which a given inference is neither allowed nor disallowed by the rules as they can be read off from the standard or core examples.

This picture of the situation is helpful as far as it goes, but it distorts the phenomenology. In ordinary life (not to be confused with life among philosophers) nobody makes a distinction between an inference which is compelling because it is dictated by the rules of inference that are in force, and an inference that is compelling because it is plausible to extend those

rules of inference into a vagueness zone, in a certain fashion. The distinction between these two sources of inferential compulsion ordinarily is not part of the phenomenology of making inferences in novel circumstances. If you pretend to be the high school algebra student, confronting the argument that the series (PS) equals the quotient $1/(1 - r)$ in every case where the division is meaningful, you are doing a poor job of pretending if you allow yourself to cleanly separate the argument into two stages: first an analogical argument with the conclusion that (roughly) infinitary applications of the division algorithm are uncontroversially appropriate, and second, an argument in which the division algorithm is thus applied to the example at hand. The student would have no basis for thinking the argument has this two-stage structure. High school algebra continually presents students with unfamiliar applications of simple algebraic methods—novel word problems that require some sort of clever parsing to get them to fit a familiar pattern, for example. Solving such problems involves an appeal to unstated auxiliary principles, such as the principle that putting more cops in a car is adding. The student cannot be expected to see that using algebraic methods to solve a problem about cops and robbers does not analogically extend the rules of the algebra game per se, whereas the argument we are considering does.

The student is confronted with an inference that seems to her "fully drawable." She runs through the inference in her mind, and she feels its logical tug. But we know she ought not draw this inference, logical tug aside. We can help ourselves to all manner of stock jargon to make this point. We can say that the student does not realize that her divisions are "merely formal." We can say that she does not realize that she lacks the security of an existence theorem, since she is innocent of existence theorems. The general point is that the student lacks a *trustworthiness criterion* for an inference that strikes her as fully drawable, that strikes her as a novel, but licit, application of the rules of a game she knows well. She lacks a semantics.

If the student took a moment to test her answer with a few values of r, she would get a shock, because it would be obvious to her that the series (PS) diverges for some values. But that would leave her in a bind: she would have a compelling argument for a conclusion she has falsified by another argument. Both of her arguments take her onto new turf, because both involve stretching to the infinite case intuitions that are good in the finite case. She has no trustworthiness criterion for either argument, and for both she needs one.

It is binds of this sort I want to discuss, as a preliminary to explaining how we cure Fred. The high school student might suffice as an illustrative example of every point I want to make. But in the case of the high school student, the trustworthiness criteria she needs are well understood by others, and she has many ways to learn that. As a result, we tend to think of her initial epistemic duty in the matter as extending no farther than asking her teacher what on earth is going on. One can argue that even for a student, there is more to her epistemic duty than that, but I prefer instead to have an example in which the person playing the role of the student does not have such an easy time of it. So I will describe a person in a very similar bind to the bind the student is in, but who lives at the very dawn of scholarly inquiry into the subject of infinite series and their generating functions. If he wants a trustworthiness criterion, he is obliged to supply his own.

There was a great and now dead mathematician and philosopher named Leibniz*, who lived in the late seventeenth and early eighteenth century. Leibniz*, who replaces Leibniz in the world I am describing, did many of the same things Leibniz did in the real world. He differed from Leibniz in two respects: first, when—in my story—Leibniz* confronts a certain conceptual puzzle, he is able to provide a semantic rule that sorts things out in a way that dissolves the puzzle. In order to accomplish that, he has to invent some mathematical theory nobody could have invented in the early eighteenth century. (People did invent it finally, but it took them a long time.) Second, according to the story I will tell, we know exactly the arguments Leibniz* toys with, and we know the order in which he toys with them. It would be possible to tell a story about the real Leibniz instead, and with heavy use of conjecture make the same points. But it is much more convenient to let Leibniz* be the main character, because we needn't argue about what he knew and when he knew it, and because Leibniz* is able to pose a problem at the beginning of research in real analysis, and then solve it by nineteenth-century methods, all inside one skull and one lifetime.

Here is the story. Like the high school student, Leibniz* starts thinking about a puzzling infinite series (he calls it an infinite sum). The series is

$$s = 1-1+1-1+1-1+ \text{(etc.)}^1$$

The series s does not make steady progress toward a "value at infinity," as does the series $1/2 + 1/4 + 1/8 + \ldots$, for example. Neither does s run off beyond every candidate value, as does $1+2+3+\ldots$, for example.

But even though it does not run off beyond all limit, s seems not to add up to any single number as its "value"; it wobbles back and forth between one and zero as it cranks along. Leibniz* discovers an argument that appears to show what the infinite sum s "is":

Argument L:

When one multiplies through the expression $(1-1+1-1+\ldots)$ by the quantity -1, the result is the quantity $(-1+1-1+1-\ldots)$. That is,

$$-1(1-1+1-1+\ldots) = -1+1-1+1-\ldots$$

So we have the identities

$$\begin{aligned} 1-1+1-1+\ldots &= 1+(-1+1-1+\ldots) \\ &= 1+(-1(1-1+1-\ldots)) \\ &= 1-(1-1+1-\ldots). \end{aligned}$$

Therefore:

$$1-1+1-1+\ldots = 1-(1-1+1-\ldots)$$

or, $s = 1 - s$. So s equals $1/2$.

It is perfectly clear to Leibniz* that every step of Argument L would be unimpeachable if the sum s were finite. For example, one can "multiply through" by -1, switching signs all along, and the resulting finite sum will add up to the result of multiplying the original sum's value by -1. Is it legitimate to make these same inferences when one is manipulating infinite sums?

Leibniz* confronts this type of question all the time. Here is a different example: when, intuitively, infinite sum s_1 and infinite sum s_2 each has a value, it seems one can add or subtract the sums term-by-term, just as one can do in finite algebra. For example, the infinite sum $s_1 = 1/2 + 1/4 + 1/8 + \ldots$ has the value 1, intuitively, and the infinite sum $s_2 = 1/4 + 1/8 + 1/16 + \ldots$ has the value $1/2$. The result of subtracting s_2 from s_1 term by term is s_2 . Happily, the result of subtracting $1/2$ from 1 is $1/2$, so in this case the subtract-term-by-term formula of finite algebra gives the intuitively right answer. But when an infinite sum runs away beyond all limit, can one still add the sum to, or subtract it from, another infinite sum term-by-term? For instance, can one add the infinite sum: $1+2+3+\ldots$ to itself term-by-term, giving the answer: $2+4+6+\ldots$? Here one cannot check the result in the same way.

It seems very plausible to say that the answer is yes, any infinite sums can be added (subtracted) term-by-term, simply because they are sums and that's what addition (subtraction) is like. It seems to be built into the concepts of addition and subtraction that this is how they work. To put it a little differently, if one asks whether it is plausible to analogically extend the add-term-by-term rule of inference from finite sums to infinite sums, the right answer seems to be "of course it is plausible; otherwise either it really wouldn't be adding at all, or the things being added wouldn't be sums at all." (Leibniz* was coming to understand that they aren't sums at all, they are sequences of partial sums. But that represents a stage of semantic self-awareness just a bit farther along than I want anybody in my story to have, for the moment.) The point I want to emphasize here is that for someone in the process of developing a theory of convergence for the very first time, transferring the inferential practices of traditional finite algebra to the case of infinite sums would not feel like a process of "redefinition" or a process of "stipulating new meanings." Something would have to alert the thinker that this is not an "ordinary" case of applying stock inferential methods to slightly novel subject-matter.

Since he is no high school student, something does alert Leibniz* when he contemplates Argument L. Certainly each step of inference is absolutely compelling. ("Of *course* you can multiply through. Otherwise this wouldn't *be* addition and subtraction and multiplication.") Still, Leibniz* cannot see how the sum s manages to add up to any particular number, given how it flip-flops back and forth between zero and one as you keep on doing the requisite additions and subtractions. As a result, he can't see why algebraic operations on the sum s are trustworthy—algebraic operations being in their nature manipulations of numbers.[2]

I want to think of this doubt as metalinguistic, as a doubt about semantic characterization. (I'll say a bit more about this description of the doubt shortly.) In metalinguistic terms, the doubt can be spelled out this way: the inferences of Argument L are compelling. So one supposes, prima facie, that they are valid. But it seems plausible that the denotation of an infinite-sum expression is the number, if any, that the partial sums "approach" as one extends the sum. It seems, therefore, that the expression "$1-1+1-1+\ldots$" does not denote any number. Now, the rules of ordinary algebra are valid because they correspond to properties of, and relations among, numbers. How can they be validly applicable to the sum s if the expression "$1-1+1-1+\ldots$" does not denote any number?

It will be useful to have a summary statement of this problem:

(I) Leibniz* has come upon a use of language which is prima facie inferentially correct but prima facie semantically unintelligible.

A second problem shows up on the heels of this one. Here is another argument:

Argument G:
Group terms in s this way: $(1-1) + (1-1) + (1-1) + \ldots$, and straightaway one sees that the sum is 0.

Ordinary finite algebra permits arbitrary grouping of terms in a sum, so Argument G, like Argument L, is very compelling. But the conclusion of Argument G contradicts the conclusion of Argument L. In summary:

(II) Leibniz* has come upon a use of language which is prima facie inferentially correct but paradoxical.[3]

Most of the time people lack semantic self-awareness, not because it is hard to have it, but because there is no reason to have it. Leibniz* has a reason. Something is going badly wrong in his inferential practice, as is clear from (II). But he cannot see what exactly is going wrong, as is clear from (I). That means he can't tell which analogical extension of finite algebraic reasoning is trustworthy when the topic is infinite sums. All of the usual forms of reasoning are equally compelling in the infinite case, and all are very compelling indeed. He needs a principled way to tell the wheat from the chaff, since his logical intuitions are telling him it's all wheat.

Leibniz* would not speak in explicitly semantic terms, because most people do not, and because even theoreticians seldom did in his time. We will do well, though, to borrow Carnap's notion of the "material mode" and think of some of the things Leibniz* says as semantic statements in the material mode, all of which could be rephrased as semantics statements in the formal mode. For example, Leibniz* says he doesn't see how the infinite sum s "can be" any particular number, because of its flip-floppy behavior. That is a material-mode semantic use of modal language; a formal mode paraphrase might be "The only account I have of why an infinite-sum expression denotes a particular number posits that an infinite-sum expression denotes the number, if any, to which its partial sums converge, and it appears the partial sums of s do not converge." And Leibniz* says that he does not see how the infinite sum s or any other "divergent" infinite sum "can obey" the laws of algebra, since no such sum is a particular number. Nor does he see how it can be exempt from those laws, since it

is after all just a sum of numbers. Leibniz*'s use of "is" in the locution "such-and-such an infinite sum is the number n" should be thought of as a material-mode form of the semantic locution "such-and-such an infinite-sum expression denotes such-and-such a number."

If we are careful to think of these "essentialist" turns of phrase as semantic, we will not be puzzled by what Leibniz* does next. What he does next (and what in reality Leibniz and a string of other mathematicians who followed him did next) is develop a new conception of what it is for an infinite sum to "be" a given number. This new conception enables Leibniz* to argue, independently of Argument L, that s is 1/2. Assuming this new conception of what an infinite sum is, Leibniz* is able to show that some algebraic manipulations of infinite sums should be trusted, whereas others should not be trusted. The algebraic inferences comprising Argument L should be trusted, it turns out, and the algebraic inferences comprising the short and snappy Argument G should not. I will translate all of these thoughts into explicit formal-mode semantic language. In my story Leibniz* will give a revised semantics for infinite-sum reasoning, a semantics which specifies a relation an infinite-sum expression must bear to some number in order for the expression to denote the number. It will be a consequence of this semantics that the inferences comprising Argument L are valid, the inferences comprising Argument G invalid, and the sentence "s = 1/2" true.

First step of Leibniz's semantic program.* Leibniz* decides that despite surface appearances, the syntax of an infinite-sum expression is not an "infinite version" of the syntax of finite-sum expressions. For example, the expression "1/2 + 1/4 + 1/8 + . . ." really has the syntax of the infinite-sequence expression

$$< 1/2, 1/2+1/4, 1/2+1/4+1/8, \ldots > ,$$

and the infinite-sum expression "1−1+1−1+ . . ." really has the syntax of the infinite-sequence expression

$$< 1, 1-1, 1-1+1, \ldots >.$$

This is an "invisible syntax" claim of the sort often made as an adjunct to semantic theory, as when one says the "true grammar" of the generalization "All dogs bark" is "For any entity, if it is a dog, then it barks." Probably one should think of these claims as belonging with "translational" semantic theories, where a portion of natural language is mapped into a formal language, for which the semantics is then given. One could

do without an invisible-syntax preliminary to the rest of the semantics, but it greatly simplifies, indeed trivializes, the task of showing Argument G invalid.

Argument G turns on "grouping" terms in the fashion

$$(1-1)+(1-1)+(1-1)+\ldots.$$

Since the infinite-sum expression "1−1+1−1+ . . ." actually has the grammatical form of

$$<1, 1-1, (1-1)+1, ((1-1)+1)-1, \ldots>,$$

whereas the infinite-sum expression "(1−1)+(1−1)+(1−1)+ . . ." actually has the grammatical form of

$$<(1-1), (1-1)+(1-1), ((1-1)+(1-1))+(1-1), \ldots>,$$

Argument G simply *miswrites* the infinite sum s, although the miswriting is invisible miswriting. Argument G is invalid. It remains to figure out whether Argument L is valid or invalid. Now Leibniz* must do some heavy lifting.

Second step of Leibniz's semantic program:* Leibniz* defines a certain relation "CS" which sometimes obtains between an infinite-sum expression and a number. If an infinite-sum expression bears the CS relation to a number, then the infinite-sum expression denotes that number. This is Leibniz*'s definition of the relation:

Relation CS (for "Cesaro-summable to"): Suppose s# is an infinite sum. Let a_0, a_1, a_2, and so forth, be the terms of s#. Let the partial sums of s# be denoted as follows: $s\#_0 = a_0$, $s\#_1 = a_0 + a_1$, $s\#_2 = a_0 + a_1 + a_2$, and so forth. Now consider numbers given by the quotient E_n:

$$E_n = (s\#_0 + s\#_1 + s\#_2 + \ldots + s\#_n) \,/\, (n+1),$$

and consider what happens to the sequence of numbers E_n as n grows large. It may be that there will be a number b such that as n grows larger and larger, the sequence of numbers E_n gets arbitrarily close to b. If that happens, we will say that the infinite sum s# bears the CS relation to the number b. If no such number b exists, we will say the infinite sum s# does not bear CS to anything.

Intuitively, this criterion for "what the infinite sum is" asks us to consider the long-term average (arithmetic mean) of the partial sums. If those average values become progressively closer to a given number as progressively longer-term averages are considered, then we call that number "the

value of" the infinite sum. In the case of the sum $1-1+1-1+\ldots$, the string of averages looks like this:

$$1/(0+1) = 1$$

$$[1+(1-1)]/(1+1) = 1/2$$

$$[1+(1-1)+(1-1+1)]/(2+1) = 2/3$$

and so on.

If you take the time to carry the process a half-dozen steps farther along, you will begin to suspect that these average values are getting progressively nearer the number $1/2$. Leibniz* suspects this, as did Leibniz. Leibniz* tinkers for a while with this idea, looking at other "alternating" infinite sums as well as s. He decides he has got a very plausible conception of "the number to which a sequence of partial sums converges"—that is, the number an infinite-sum expression denotes.[4] He does, too, although the mainstream mathematical tradition has preferred the more fruitful idea that a real-number series converges to whichever number is the limit of the sequence of partial sums, *simpliciter*. (That was essentially the idea Leibniz* began with in his intuitive thinking about the sum s.)

Third step of Leibniz's semantic program:* eventually Leibniz* works out a full theory of this long-term-average concept of convergence, and draws some conclusions within that theory.[5] We can safely omit the details (which get a bit detailed even as details go) because the work really was done eventually, mainly by Cesaro and Hardy. Here are the conclusions Leibniz* draws:

(a) The infinite sum $1-1+1-1+\ldots$ bears the CS relation to the number $1/2$. So the expression "$1-1+1-1+\ldots$" denotes the number $1/2$. In material-mode talk: the sum s exists and is $1/2$.

Whenever an infinite-sum expression "Σa_n" denotes the number b, algebraic inferences corresponding to the following two principles are valid:

(b) If $\Sigma a_n = b$, then $\Sigma k a_n = kb$; where k is any number.
(c) If $a_0 + a_1 + a_2 + \ldots = b$, then $a_1 + a_2 + a_3 + \ldots = b - a_0$; and conversely.

Facts (b) and (c) show that the steps of algebraic reasoning used in Argument L are valid inferences (as the reader may verify easily). Leibniz* has finally arrived at a highly specialized semantics for a certain portion of

mathematical language. His semantics has a special-purpose criterion of argument validity—valid arguments involving the application of algebraic methods to infinite sums are exactly the arguments the leading principles of which may be derived from the "CS criterion" of infinite-sum convergence. Otherwise, we may suppose the semantics is classical, operating with the concepts of denotation, truth, and truth-of, with validity in general understood as truth-preservation.

Let's take stock. Leibniz* found himself confronted with some inferences he found intuitively compelling. They were algebraic inferences, though applied to infinite sums rather than finite sums. It seemed plausible to analogically extend traditional algebraic inferential practice to this new case, because infinite sums appear to have almost exactly the same grammatical "shape" as finite sums. It seemed as though the only vocabulary involved in infinite sums is the vocabulary of traditional algebra—numerals, plus and minus signs, and so forth, concatenated in very much the usual forms. The traditional algebraic inferential practice just is the practice of inferentially manipulating this vocabulary in certain allowed ways. We summarized this by saying that Leibniz* had come upon a use of traditional algebraic language that was prima facie inferentially correct. So far, he was much in the same position as the high school student confronted by her "argument from division."

And, again like the high school student, Leibniz* found intuitively compelling reasons to think his first line of thought had gone wrong. He could not see—at least he could not see at first—how the infinite-sum expression $1-1+1-1+\ldots$ could denote a number, and it seemed to him that algebraic inferences could be valid only if applied to numbers, or to "things that are numbers." So the analogically extended inferential practices he found so compelling were not semantically intelligible. Worse yet, it was possible to find a pair of equally intuitively compelling inferences with contradicting conclusions. So the analogically extended inferential practices were paradoxical. Here again are the slogans from earlier:

(I) Leibniz* has come upon a use of language which is (for him) prima facie inferentially correct but (for him) prima facie semantically unintelligible.
(II) Leibniz* has come upon a use of language which is (for him) prima facie inferentially correct but paradoxical.

I claim that once Leibniz* realizes that (I) and (II) are true, we should hold that he has no right to produce any of these "infinitary" algebraic ar-

guments as reasons for a belief about alternating series. (This localizes the problem maximally—though one might argue that it really is a problem touching on all infinite series, not just alternating ones.) *He has become semantically self-aware,* and his semantic self-awareness includes recognizing that some or all arguments of a certain class may be invalid even though they seem compelling. Some of those arguments may be valid, too, but he cannot tell which is which. It is illogical of him to put forth any one of these arguments *as giving a reason* until things have been cleared up at the level of metatheory. It matters not that he would not, cannot, express his metatheoretic knowledge in an explicit metalanguage. Most people never do, but that does not prevent them having beliefs and knowledge concerning the semantics of their languages. Leibniz* can say that he does not yet know whether the sum $1-1+1-1+\ldots$ *really exists* instead of saying that he does not yet know whether the expression $1-1+1-1+\ldots$ *denotes* anything.

Eventually, when Leibniz* has invented his new semantics, and with it a new syntax or grammar, (I) and (II) will cease to hold. He will have a clear rationale for trusting some algebraic manipulations of infinite sums—for instance the manipulations driving Argument L—and a clear rationale for not trusting others—for instance the manipulations driving Argument G. Now he may put forward, as reasons for a belief, such arguments as his new-found semantic appreciation ratifies as valid. Here, too, the fact—if it still is a fact—that these arguments have a "felt compulsoriness" about them is irrelevant. Once a thinker's blissful semantic naiveté has been compromised by her coming to have semantic self-awarenesses parallel to (I) and (II), there is no going back to the garden. Until the thinker finds, or is handed, a semantic ratification of some of the analogically extended inferential practices for which she has come to realize problems (I) and (II) exist, she has no right to deploy any of those practices in her inquiries, to generate reasons. Of course she need not speak of these moments of her thought in terms of "denotation" and "validity"; she may, and ordinarily will, speak in the material mode.

15

Two Charleys

Now we have in place the requisite machinery to understand how Fred's confusion can be "cured." Our account of his cure must explain what happens when Fred "discovers" that there are two big ants, not one, and that he has been mixing them up. It is easy for our account to explain the last bit of the scenario—a "cured" Fred realizing that he has been mixing up the two big ants. That is just a self-ascription by Fred of "former confusion," exactly like our self-ascription of confusion when we recall the events in the parking lot. We need to think harder about the rest of the scenario. Let's start by elaborating the story of Fred and his ants to include a "discovery" by Fred of a second big ant.

Suppose Ant A is up top in the ant colony drinking some water, while Ant B is down below sleeping. We give Fred a fiber-optical device enabling him to see both ants at once. This twist in the story is aimed at keeping a little control over what elements have been changed: if Fred sees both ants at once without much altering the normal flow of events in the ant colony, it is likely that any ensuing changes in our epistemic intuitions are reactions to changes in Fred's epistemic states, with very little admixture of reactions to changes in the object of Fred's epistemic states. Fred glances back and forth between the top of the ant colony and its innards, sees both big ants at once, and is astounded.

It might happen that Fred instantly alters his use of words. He might straightaway say something like: "I'll be damned. There are two Charleys. One Charley is down below asleep, while the other Charley is up above drinking." He might remain faithful to this new way of talking from that moment on. It is a new way of talking only in the sense that it is a new way for Fred to talk about events in his ant colony, but it is familiar English usage. One can use a proper name as a count-noun. As standard English usage, this way of talking carries with it an inferential practice, also stan-

dard and familiar. Fred can simply take up this practice, constructing Argument A:

(Premise) One Charley is down below sleeping; another Charley is up top drinking.
(Premise) The first Charley will wake up soon.
(Premise) When the first Charley wakes up, the second Charley will get ready to go down for a nap.
So,
(Conclusion) Soon the second Charley will get ready to go down for a nap.

Every one of us could adopt this way of talking and reasoning about Fred's ant colony, with ease. As a first approximation to the grammar of the argument, the expressions "the first Charley" and "the second Charley" anaphorize back to the sorted quantifiers "one Charley" and "another Charley," so that the anaphors make sense in the context established by these quantifiers. The argument has got some dependence on tense and temporal particles, but except for that complication there is no reason not to treat "Charley" as a count-noun throughout, with a semantics like the one proposed by Anil Gupta.[1]

If Fred adopted this new way of using the word "Charley," complete with the grammar and the inferential practice normally associated with that use of "names," he would have become a new man. Nothing to "forgive." There would be no reason to apply to Fred a semantics different from the semantics we would apply to ourselves when we engage in two-Charley talk and make the associated inferences. Fred no longer would deserve inferential charity. Therefore he would no longer be confused.

Fred could go down other linguistic paths, spurning the "two Charleys" option. He could name the two big ants "Ant A" and "Ant B" (to our amusement). Again no more grounds for inferential charity—surely the most plausible semantics for us to apply to his reasoning would be classical, with both "Ant A" and "Ant B" interpreted as singly Referring terms. Again Fred would be cured.

Or, Fred could continue to use the single name "Charley," but with studied ambiguity. No special charitable semantic treatment would be in order then, either. Again Fred would no longer be confused.

What is striking, though, is that Fred would not have to adopt any "new language" in order for us to think it reasonable to change our opinion that he was confused; that is, in order for us to change our semantic position.

Suppose Fred starts to talk about his startling observations. "This is quite extraordinary," he says. "It appears that Charley is up top drinking and down below sleeping at the same time. Makes no sense." Imagine that he quickly makes clear that he does not believe he is suffering an illusion (although he might dawdle for moment, questioning the reliability of our fiber-optics, or wondering whether some clever practical joke is being played). At this juncture, before Fred switches to a new way of talking, we would—should—stop granting him a general right to such arguments as Argument B:

I have watched Charley carry large twigs.
I have watched Charley carry large pebbles.
Therefore,
Charley can carry large twigs and he can carry large pebbles.
(So Charley is very strong, etc.)

If Fred were to give this argument, more than one semantics would be appropriate, but none of them would be inferentially charitable. We might, for instance, apply an "ambiguity" semantics, allowing the possibility that the tokens of "Charley" in the two premises denote distinct ants, invalidating the argument. We would require that some criterion for coreference be met before we were ready to count this token of its argument-type as valid. Fred, we would be tacitly insisting, *can* distinguish Ant A from Ant B in thought—a semantic judgment, another semantic use of psychological language, in this case modalized. Again, Fred would be cured.

What would happen if, instead of registering profound bewilderment at what he simultaneously saw through the fiber-optic device and through his unaided eyes, Fred merely shrugged it off and went on as usual? Probably we would search for psychological explanations of his evident inability to take in what he was seeing (emotional attachment to "his pal Charley," or whatever), but—so my intuitions tell me—we would not revise our policy of inferential charity. We would continue to say that Fred cannot distinguish Ant A from Ant B in thought (a semantic judgment) even though he has been shown both big ants at once.[2]

It is easy to understand why we would alter our semantic appraisal of Fred's reasoning if he switched to a new ant-colony language—"two Charleys," or whatever. It is less easy to understand why we would alter our semantic appraisal merely because he looked through the instrument

and immediately registered bewilderment. I suggest that the correct explanation is that Fred's bewilderment is at root semantic. The situation Fred finds himself in when he looks through the fiber-optic device exactly parallels the situation Leibniz* found himself in when he contemplated Argument L and Argument G. The lines are drawn less clearly in Fred's case; unlike Leibniz*, Fred is not a theoretician by trade and does not introduce radical new patterns of thought which he then is obliged to justify, explicitly and publicly, to the extent possible. Fred can look through the device, say "Damn," and go about his business. We could only infer the nature of his vexation. I have told the story in such a way that Fred gives us just enough hints for us to be sure what is troubling him, even including the comment "Makes no sense." Fred has come upon a strange apparition, an experience that "makes no sense." But he has also come upon a use of language that makes no sense, and that is why the experience makes no sense. He is confronting the prima facie correctness of Argument C:

Charley is (now) up top drinking.
Charley is (now) down below sleeping.
So:
An ant can be up above and down below at once.

Worse yet, he cannot see why he should not accept the premises. Something has to go. Maybe he shouldn't accept the premises, or at least one of them. Maybe the conclusion doesn't follow. But if it doesn't follow, why doesn't it? Fred is not baffled about the behavior of ants; Fred is baffled about the validity of inference.

We can reiterate the diagnosis of Leibniz*'s situation almost unchanged:

(I_f) Fred has come upon a use of language which is (for him) prima facie inferentially correct but (for him) prima facie semantically unintelligible.
(II_f) Fred has come upon a use of language which is (for him) prima facie inferentially correct but paradoxical.

In Fred's case, a single argument—Argument C—does the work of both Argument L and Argument G in Leibniz*'s case. Argument C seems a good inference (seems so to Fred), but it leads to paradox. Once Fred realizes that (I_f) and (II_f) are true, we should hold that he has no right to produce *any* of the "Charley" arguments he has been fond of giving. Fred has

become semantically self-aware, and his semantic self-awareness includes recognizing that some or all arguments of a certain class may be invalid even though they seem compelling. Fred would not put it that way—of course—but it is a fair summary, in spruced-up philosopher's lingo, of his state of mind. As a result, it is illogical of Fred to put forth any "Charley" arguments as giving a reason until things have been cleared up at the level of metatheory, though the metatheory can be formulated in the material mode.

Since it is illogical of Fred to put forth arguments which, a few minutes or seconds earlier, it would have been entirely logical of him to put forth, we are obliged to make compensatory changes in our semantic position concerning how to appraise his arguments for validity. We no longer should be inferentially charitable. We should (and—it seems intuitively—we would) switch from the judgment "Fred cannot discriminate between Ant A and Ant B in thought" to its contrary, "Fred *can* discriminate between Ant A and Ant B in thought." If conditions (I_f) and (II_f) do not ensue, for example because Fred is prevented by psychological factors from reaching even this minimal level of "negative" semantic self-awareness, we have no obligation to modify our charitable semantic position, because we have no basis for thinking Fred's "Charley" arguments any less logical than we have previously thought them.

Formal logic, as a tool for evaluating the "logicality" of someone's reasoning, tends to deflect attention from the component of logicality emphasized here and emphasized in the case of Leibniz*. This happens because we are accustomed to doing logic for settled inferential practices where there is wide agreement on what ought to be inferred from what, at least in simple cases. Neither Fred nor Leibniz* knows exactly how he should infer even simple things (about a circumscribed subject-matter); or rather, neither of them knows *at the start of his inquiry* how he should infer things—counting Fred as starting off on a new inquiry the moment he looks down the fiber-optics device. For both Fred and Leibniz*, semantic metatheory, and the "object theory" of which it is the metatheory, develop hand in hand. These are not rare events in the history of thought, especially the history of science. But they are rare in mundane life.

An exception to this rule occurs when a confused person comes to have experiences that "reveal to her that she is confused." Inevitably, this obliges the person to make radical changes in her inferential practices. Until the new inferential practices start to fall into place, the person should

recognize the extent to which her extant inferential practices do not make sense. This process—consciously losing semantic grip on one's standing inferential practice, getting some idea what has gone awry, and replacing the practice with a new and semantically justifiable one—*is* the "revelation of distinctness" engendered by the confused person's liberating experience.

16

Young Newton

Now and then someone claims to have "made logically visible" a distinction nobody else has known how to see. In effect, the person who makes such a claim is charging others with confusion, with having always confused the objects she, the insightful one, has learned are distinct. We can apply the lessons learned in Chapters 14 and 15 to these cases. In particular, we can apply the idea underlying Fred's "cure."

When a person comes upon a use of language with an associated inferential practice, but this practice is (at the moment) both semantically unintelligible and paradoxical, the person has no right to rely on this new inferential practice as providing reasons, until and unless she finds a way to semantically characterize the practice and tell wheat from chaff.

Fred "comes upon" such a use of language in himself, but the general idea is wider. Therefore we should be alert to the possibility that a debate between person (or group) A, purporting to have made logically visible a distinction not previously logically visible, and person (or group) B that rejects this claim, may take the following form: B makes claims exactly like claims (I_f) and (II_f), except with A as the subject; then B follows with the charge that in the presence of these "facts," A has no right to deploy any of her newfound types of argument as reasons for anything. A ought then to reply by giving a semantic characterization. If she fails to do so, her critic has won the round.[1]

I will discuss, as an illustration, just one example of a debate that seems to me to have had precisely this form: Bishop Berkeley's attack on analysis. It is a simple story to tell, and I will make it even simpler by letting the young Newton, the Newton of the 1666 manuscripts, play the role of Berkeley's target. He was not, for these writings were not published until long after Berkeley launched his attacks. Moreover, the leading idea of Newton's technique evolved slowly through the work of many others—

not least Newton's mentor, Barrow. Finally, not all of Berkeley's points can be made against the 1666 manuscripts. But Newton's presentation was crisp and uncluttered, which for our purposes helps immensely. And nothing will be lost if we pretend it all just came to young Newton in a flash.[2]

Young Newton considered two moving bodies, A and B. Assume neither is moving with a uniform velocity. During some arbitrary period of time, body A travels distance x, and during the same period of time body B travels distance y. Newton considered what would happen if each of the bodies moved for an additional *moment* of time—an infinitely little period of time—designated "o." During that moment, it would seem, body A moves with just one velocity—call it "p"—since there is not enough time during a mere moment for the velocity of the body to change, even though over finite periods of time the velocity does change. Likewise body B moves with just one velocity q during the same moment. One can ask how far each body moves during the moment o, and the answer clearly is that body A moves a distance of op, the elapsed time multiplied by the rate at which it moves during that time, while body B moves a distance oq, by symmetric reasoning. One might think of op as the momentary augmentation of the distance x resulting from the momentary movement of A after it traverses distance x, and one might think of oq as the corresponding momentary augmentation of distance y. After the moment o has elapsed, therefore, distance x has become distance x + op, while distance y has become distance y + oq.

The distances x and x + op are not the same, and the distances y and y + oq are not the same. A person might find this hard to accept. After all, one cannot *see* any difference between, for instance, the total distance traversed by body A in five seconds and the total distance it will have traversed after a mere "moment" more has elapsed. Nor can one measure the difference by laying down a ruler. It will not help to use a magnifying lens. Or the finest microscope in Europe. Or any *possible* microscope. Tough, says the young Newton. There are two distances, x and x + op, not one distance. If he is right, people have been confusing two distinct distances or lengths, just as Fred confused Ant A and Ant B. They have been in even worse shape than Fred, though, because nobody *could* hand them a fiber-optic device that would finally enable them to discriminate perceptually between x and x + op.

How did the young Newton detect this difference that no possible microscope could reveal? Simply by employing his new linguistic practice of

moment-talk, together with its associated inferential practice. (Remember, we're oversimplifying the history.) In this new inferential practice, x and x + op (or y and y + oq) behave differently. Substituting x + op into an equation is not the same as substituting x into the equation. Dividing by x + op is not the same as dividing by x. And so forth. But there were additional forms of inference—equally plausible, Newton believed, that have no counterpart in the algebra of ordinary finite quantities. Occasionally he would divide some quantity by a moment. For example, he might divide the quantity $aboox^2 + b^2ox$ by the moment o, getting $abox^2 + b^2x$. One does not divide by zero in ordinary algebra, so a moment had better not be zero. But in the next line of his argument he might well equate the result of his computation, the quantity $abox^2 + b^2x$, with the quantity b^2x, with the comment "those terms are infinitely little in which o is," so they count for nothing in comparison with terms "in which o is not." That makes a moment seem rather like zero after all.

A skeptical contemporary of young Newton might have said: "it is unclear *what a moment is.*" This is another example of material-mode semantic language. If we were to take it at its essentialist face value, we never would figure out what the logic should be of a debate in which such language occurs, since we do not know how one ought to reason, in general, about "what a thing is." But if we accept that the language has a semantic point, we have a hope of understanding a debate in which it occurs—as a debate about semantic policy. So we should understand a question about "what a moment is" as a question about what, if anything, a moment expression denotes; or more generally, what the semantics of moment-reasoning is.

Seen in this light, young Newton was in the same box as Leibniz*. Like Leibniz*, he wanted to use novel forms of inference which he could not fully explain or understand semantically. Moreover, some of the arguments typical of this new form of reasoning conflicted, generating "paradox," although in Newton's case the conflict was really only a puzzle about how moment expressions could have a coherent semantics—for example, how they could denote zero and not denote zero, as it seems on the semantic face of things they must. This is different from the paradox Leibniz* confronted, which arose entirely in the object language, as it were. So we cannot repeat the conditions (I) and (II) that applied to Leibniz*, or the strictly analogous conditions (I_f) and (II_f) that applied to Fred, without some stretching. But the gist of the idea remains.

Newton has come upon (in his case, invented) a use of language that seems inferentially correct. In it one multiplies and divides quantities, one makes substitutions of equals for equals, and in general one does the sorts of thing one knows how to do from traditional algebraic practice. But some of the quantities are conceptually novel, just as some of Leibniz*'s quantities were conceptually novel. So there is a question what inferences in this new inferential practice are valid, even though there is a felt naturalness to all of them. And as with Leibniz* and Fred, the question is not idle, as one can get conflicting answers according to what one really means. Just as with Leibniz*, and just as with Fred, it seems Newton has no right to use any of these forms of inference, pending semantic ratification.

Berkeley's criticisms should be seen as making exactly these points. In fact this meaning is pretty much on the face of Berkeley's texts, though it is not the meaning those texts usually are given.

Berkeley allowed that it was easy to follow the Newtonian inferences in a purely "formal" way—one knows how to make the moves—but often those inferences made no sense given what one takes the language of geometry and algebra to signify (Berkeley's reference to "fluxions"—a later notion in Newton's thought—can be replaced without loss of Berkeley's meaning by "moment," or "momentarily augmented distance," or some other expression more in tune with young Newton's way of talking).

> . . . Men too often impose on themselves and others, as if they conceived and understood things expressed by Signs, when in truth they have no Idea, save only of the very Signs themselves. And there are some grounds to apprehend that this may be the present case.[3]

> Nothing is easier than to assign Names, Signs, or Expressions to these Fluxions, and it is not difficult to compute and operate by means of such Signs. But it will be found much more difficult to omit the Signs and yet retain in our Minds the things, which we suppose to be signified by them.[4]

Berkeley thought of "Ideas" as images, with visual images the paradigm. Therefore one might suppose at first glance that his point in these passages is just that you cannot picture the difference between a line and that same line "momentarily augmented." But that is not at all his point. He means to contrast the intuitive smoothness of inferring things ("computing and

operating") in Newton's framework with the difficulty of figuring out ("retaining in our Minds") what is "signified by" the things you are inferring to and inferring from.[5] It is like a sailor

> who may practically apply certain Rules derived from Astronomy and Geometry, the principles whereof he does not understand: And as any ordinary Man may solve divers numerical Questions, by the vulgar Rules and Operations of Arithmetic, which he performs and applies without knowing the Reasons of them: Even so it cannot be denied that you may apply the Rules of the fluxionary Method.[6]

Newton needs to reason about the ratios of momentary, seemingly punctile things, things without magnitude, but "ratio" only applies to what has magnitude; Newton needs to consider the first (or last) moment of a moment itself (for reasons we have not considered), but a moment is not the sort of thing that has a first or last moment ("he who can conceive the beginning of a beginning, or the end of an end, . . . may perhaps be sharpsighted enough to conceive these things").[7] If Newton really is talking about what it seems he is talking about, what he says sometimes just makes no sense. Ratios are not the sort of thing that can play the role here assigned them, nor are moments the sort of thing that can have a beginning or end. It just is not clear at all what Newton's words actually "signify," however fluently we all can speak his language and follow his arguments. Elsewhere in *the Analyst* this line of thought shades into a complaint that Newton is flatly contradicting himself; for example, by all that dividing by a moment and then claiming it annihilates quantities it multiplies, just as zero does.

Berkeley wants to be told something coherent about the signification of the language in which these smoothly natural inferences of Newton's are formulated, and he wants that semantic account to justify relying on the inferences, or at least tell us when we can and when we cannot so rely. He wants us to hold ourselves to a higher standard than the sailor's. This is very much the stance we took toward Fred when he was—if only momentarily—semantically baffled about his use of "Charley" language after he saw both big ants at once. But Newton—like Leibniz* but unlike Fred—has got himself into this situation by being linguistically creative. Fred can go ahead and start saying "the first Charley" and "the second Charley," and we will no longer doubt that he has a semantic grip on his inferences, at least as good a grip as the rest of us have. We are confident we know what it is for ants to be distinct, and we are confident Ant A and Ant B are

distinct; it remains only for Fred to start talking about them accordingly and do it in a semantically intelligible language. As soon as he does start talking that way, we'll be able to spot it, we think, and then we once again can grant him a right to rely on his ant-colony reasoning. By contrast, Berkeley did not have the faintest idea how to semantically ratify Newton's inferences, nor did anybody else. Nor did Berkeley or anybody else know of an ordinary and familiar way to talk and reason about the "distinct objects" x and x + op.

Today, when we have a surfeit of proposals for talking about infinitely little quantities, it is possible to make semantic sense of Newton's language and tell the valid from the invalid among his inferences. But there is a residual problem, unlike anything that comes up in the course of our interactions with Fred. Had Berkeley been able to appreciate this problem, he would have had a second tier of criticism.

Suppose, for example, that we construct infinitesimal quantities—perhaps understood as infinitesimal bits of space—after the fashion of Lawvere, Dubuc, and "synthetic differential geometry."[8] It is well understood how to give valid arguments in which infinitesimals, so construed, are denoted. One then need only make "fair enough" translations of Newton's talk into this talk, and one has in effect a semantics for Newton's talk. But certain principles of classical logic are abandoned in the semantics; for example, the principle that a proposition follows from the negation of its negation.[9] Imagine, if you will, going up to Berkeley and confronting him with essentially this argument:

> Bishop, you are perfectly right to think that any quantity which is diverse from zero is some ordinary number, with the properties of ordinary numbers. Since a moment o is not an ordinary number, it is not diverse from zero. But you fail to see that from the premise that o is not diverse from zero, it does not follow that o is equal to zero, since that is an instance of the form of argument $\sim\sim p \rightarrow p$, and in order to ratify a good deal of Mr. Newton's reasoning as valid we must reject that principle.

Berkeley, being no fool, would inch toward the door, trying not to alarm us for fear of a more dangerous outburst of lunacy. "I see," he would say; "the opposite of Diversity is not Identity. Well done. I have to be going now."

It seems right that we have made semantic sense of Newton's infinitesimal reasoning in part by altering the concepts of sameness and differ-

ence. One can tinker with various classical logical principles without having the queasy feeling that sameness is no longer quite the same thing, but if inferences by double negation are no longer valid, it does seem that a crucial link between sameness and difference has been broken. Other links remain, so some concept of sameness and some concept of difference remain as well.[10] Berkeley's second-tier argument against Newton could go like this:

> At the moment you are no better off than the sailor. But these loonies from the late twentieth century maintain that much of your reasoning can be semantically ratified by interpreting your "moment" talk as referring to certain objects they know how to construct. But they freely confess that they can accomplish this only by applying unusual logical standards to your reasoning, standards that alter the very concepts of sameness and difference. This casts doubt on the claim (admittedly never made in so many words by you) that you have made logically visible a *distinction* that previously was logically invisible; or in other words, that you have dispelled a confusion of two *distinct* things, a distance x and the momentarily augmented distance x + op. You have not shown that these things can be coherently conceived as distinct—that is, reasoned about coherently in a language that uses the expressions "x" and "x + op" to denote distinct items—except by reinterpreting what "distinctness" is.

The gist of this objection is that Newton exposes a "confusion" suffered by his predecessors only in the attenuated sense of a "confusion" of things that count as "different" according to a sense of "different" these predecessors would not have accepted as "real" difference. This is not an objection to Newton's practice of the calculus, unless some further claim is added to the effect that the concepts of sameness and difference may not be changed for any purpose. Almost certainly the real Berkeley *would* have wanted to add that further claim, but we need not pursue the matter.

Evidently I have been papering over an important subtlety by using the very simple story of Fred and his ants as our working example of confusion. Sophisticated distinction-making often is a blend of (i) arguing for the distinctness of X and Y, and (ii) introducing a new kind of sameness and difference *by arguing for* the distinctness of X and Y. In these circumstances, when one charges a person (or a civilization) with having confused X with Y, it is less natural to think of this confusion as an "error" than it is in Fred's case.

VII

FLEXIBLE SAMENESS

17

Self-Induced Confusion

Sometimes one encounters a theorist who identifies things we do not identify, and does so on purpose, recognizing that the things could be distinguished. In the early modern period in philosophy this was quite common. The scholastic practice of distinguishing among distinctions—real, rational, and so forth—was equivalent to a practice of distinguishing among samenesses, with none of the several varieties of sameness regarded as "plain, garden-variety identity." Many of the early moderns who did not consider themselves scholastics were equally flexible, sometimes identifying As with Bs in their thought, sometimes holding As and Bs distinct. As a result, they were able to move in and out of "funny ontologies," according to the demands of theory. The entities of one of these funny ontologies can be thought of as amalgams of entities we take to be distinct.

It is natural to describe the situation in the terms I just used. Someone "identifies distinct items in her thought." Or she gives arguments which "operate within a funny ontology, an ontology of amalgamated objects." These descriptions are suggestive, and suffice for many purposes. But they do not suffice when serious questions are raised about how we should appraise the statements and arguments given by someone who has elected to identify things we take to be distinct. The analogy with Fred's confusion of Ant A with Ant B is striking. It is natural to say that Fred has identified the two ants, or that he is thinking within a funny ontology, an ontology containing neither Ant A nor Ant B, but containing instead "the big ant Charley," an amalgam of Ant A and Ant B. These descriptions are also suggestive, and suffice for some purposes.

Commentators often give, or presuppose, an especially simple account of what is going on when an early modern philosopher flexes the concept of sameness here and there, identifying things we think are diverse. They just deny that any such thing happens—without really contemplating the

possibility that it does—and claim instead that the philosopher "equivocated" in her use of some term of art, or "had several senses" of the term. This is understandable. For example, in some sense of "thought," Descartes thought that "ideas" were two diverse things: mental acts and the "internal" objects of those acts. Any competent twentieth-century philosopher who thought such a thing would marshal two senses of the term "idea," not to be mixed up on pain of committing a fallacy of equivocation. But Descartes was happy to make identifications of things he knew how to distinguish, even across categories—for example, identifying matter with its principal attribute, extension. And he was happy to identify act with object when an argument called for so doing. No doubt we have the option of dismissing all of these arguments as fallacies of equivocation, as we must do if the two different things he called "ideas" are, as they are, two different things. But it seems uncharitable.

I believe the only coherent account we can give of the practice of "strategically identifying the diverse" is as self-induced confusion. When one says of a philosopher that she "identifies distinct items in her thought," or that she "gives arguments which operate within a funny ontology, an ontology of amalgamated objects," one is taking a semantic position exactly like the semantic position one takes by saying Fred thinks Ant A is Ant B. The difference is that when Fred gives arguments that blur the distinction between Ant A and Ant B, he has no choice. The philosopher has a choice, and exercises that choice. The philosopher's confusion is self-induced, purposeful.

When one says of Descartes, for instance, that at a certain place in his philosophy he identified acts of awareness and (internal) objects of awareness, one is making a "local" semantic proposal: Descartes' arguments in that sector of his thought ought not be invalidated as fallacies of equivocation, but should be given a fighting chance to turn out valid. A natural, colloquial way to put it is that in those sectors of his thought he should be seen as thinking in terms of "act-objects." Elsewhere in Descartes' philosophy he may make claims that should, logically, discriminate between awarenesses and their objects. When we go to evaluate his arguments in one of these other sectors, inferential charity ought not be invoked. If he stumbled and thought in terms of act-objects, rather than thinking in terms of acts or else thinking in terms of objects, shame on him—write off his argument as fallacious.

We should look in detail at an example. Two considerations speak in favor of choosing, as an example, the act-object identification in early mod-

ern philosophy. First, it was very widespread, maybe universal. That is, many early modern philosophers at least sometimes reasoned in ways that "identified" acts of awareness with the intramental objects, the intentional objects, of those acts. Second, the act-object identification is an identification of items across categories, by our metaphysical standards, and therefore qualifies as a category mistake. One of the morals of this story is that we can fully appreciate reasoning which is "unintelligible" in the category-mistake sense, although there are costs. Of course we already should have drawn this moral from the simple case of Fred—it is unintelligible how Ant A could be Ant B, but this does not keep us from evaluating Fred's arguments as valid, and knowing exactly what we mean by that.

18

The Theory of Ideas

The theory of ideas was first and foremost a theory of the "aboutness" of thought: people have thoughts, and some of these thoughts are awarenesses, graspings, or representings of objects. I will stick to the term "awareness." An awareness of X is a relational complex in which a person, or perhaps the person's mind, bears the aware-of relation to X. The object X might be anything at all—cold, the sun, oneself, the property ascribed to all triangles by Euclid 1.32. It is easy to distinguish the relational complex, the object-directed state of mind, from the object X. It is easy for us to make that distinction, and it was easy for early modern philosophers to make it. But sometimes they did not make it, strategically. They chose to think of acts as objects. Here again is my suggestion: this historical judgment—"They chose to think of acts as objects"—is an attribution of confusion, the confusion of act with object. When one says "They chose to think of acts as objects," one is taking a semantic position, not ascribing a mental state to early modern philosophers. The verb "think" has a semantic use.

An important reason why we should view the statement "They chose to think of acts as objects" as a colloquial way of taking a semantic position, rather than as psychological commentary, is that we can be faced with two apparently conflicting interpretative judgments. We must find a way to square them. Consider some philosopher of whom we are disposed to say "she thought of acts as objects" (or of whom we are disposed to say "At *this* place in her writings she was thinking of acts as objects"). Here are the two apparently conflicting judgments:

(I) We want to represent the philosopher as having perfectly logical and nontrivial *reasons* for certain philosophical positions. We find that it is possible to view her in that light as long as we think of her as

identifying act and object, or reasoning in terms of act-objects. Then her arguments are logical, so that our dispute with her can be over other matters of philosophical substance.

(II) We find the philosopher's *beliefs* in some sense unintelligible, because she is making a category mistake.

The conjunction of (I) and (II) confronts us with a metaphilosophical puzzle: how can beliefs that are somehow unintelligible fit together into perfectly good reasons for holding beliefs—maybe even deep and exciting reasons? I believe that in order to untangle this puzzle we must recognize that when we make the judgment that the philosopher has chosen to think of acts as objects, we are taking the semantic position that her reasoning ought to be evaluated by means of a charitable criterion of validity, one that does not require that she respect, in her inferences, the distinction between awarenesses and the objects thereof.

Inferential charity always is a free semantic choice; it is not something one can "prove" is required. But unless we can make out at least a plausibility argument for that choice, our intuitions will abandon us as the discussion goes along. We will be unable to shake loose from the worry that any philosopher who commits a category mistake is just blundering, so that (II) above implies that the attitude expressed in (I) is wrongheaded. So let's put on the table a few reasons why the act-object identification struck early modern philosophers as both plausible and important. If we do this, we will have at the same time put on the table a few reasons why we, as interpreters of the early moderns, might plausibly elect to semantically evaluate "act-object reasoning" in a charitable spirit. My claim never will be that *we* can find the identification of act and object fully intelligible. My claim will be that we can appreciate reasoning that embodies that identification. One point of this and the following chapter is to show by example how one can appreciate reasoning that embodies an unintelligible cross-category identification.

Let's begin with what one might call "the founding analogy" of the modern theory of ideas. Call it "the worship analogy." My point here will be that the act-object identification flows naturally from the worship analogy, given a few additional metaphysical assumptions or emphases. These metaphysical emphases were prominent in the early modern period.

Descartes innovated the modern conception of the aboutness of thought by a sweeping analogical extension of the way a Christian worshipper's exhortations, pleadings, and remonstrations were taken to be

confrontations with Christ or Mary or—likeliest, perhaps—the local saint. Descartes said, by way of definitional explanation, that those thoughts that are "ideas" are, as it were, images of things.[1] "Imago," standardly an icon in a church, was a terminological choice intended to register the direct engagement of the worshipper with something not "really" in the church at all. Kneel before the statue, and you kneel before the saint; shake your fist in outrage at the statue, and you shake it at the saint. Whatever the church fathers and the great scholars—indisposed toward idolatry as they were—may have believed or wished, learning to practice popular Christianity included learning how to directly confront a dead man or woman by carrying on in front of a statue. Ordinary folks still knew how to worship the moon as Diana, except the object of worship no longer was either the moon or Diana. They knew how to entreat, beg, query, even curse the statue as the saint.

As an illustration of Descartes' analogical extension, take the case of one's thoughts about the sun. These thoughts are about the sun because they involve one's idea of the sun. As we will see, sun-thoughts involve this idea by *being* it. And one's idea of the sun, in turn, *is* the sun, existing in that manner in which things are wont to exist in the understanding (to be contrasted with the manner in which things are wont to exist in skies).[2] Here is the structure of the analogy:

[Base case of the analogy]

The worshipper
confronts, engages
the saint (who is not "materially" present to the worshipper)
by
confronting, engaging
the statue (which is "materially" present to the worshipper)
and in this way
the saint is present to the worshipper, confronted and engaged.

[Analogical extension]

the thinker
experiences, manipulates in thought
the sun (which is not materially present to the thinker)[3]
by
experiencing, manipulating in thought
a sun-thought (which is materially present to the thinker)

and in this way
the sun is present to the thinker, experienced and manipulated in thought.

For our purposes, the bottom line of the analogy can be put this way:

(1) The worshipper confronts and engages the statue as the saint. In this way the saint is present to the worshipper; the saint exists in the worshipper's acts of engagement. (Traditional lingo would have been "the saint exists in the worshipper's act of love," which strikes me as a sugared description of the gamut of attitudes the worshipper can take toward the statue/saint.)

Now, the sun is in the sky. But the thinker's sun-thought, or awareness of the sun, is right there in the thinker's mind to be manipulated (so Descartes believed, for independent reasons), just as the statue is right there in the church. So the thinker's sun-thought is the entity the thinker must manipulate in thought in order to manipulate the sun in thought. Therefore we can say, by analogy with (1), that:

(2) The thinker experiences and manipulates in thought her sun-thought (the mental act or episode) as the sun. In this way the sun is present to the thinker; the sun exists in the thinker's thinking.

One can object to (1), and we all would. The statue merely depicts the saint, we would say. It might be correct to claim that the presence of the statue is a vivid reminder of the saint, but in no way is it the very presence of the saint to the worshipper. That is a silly confusion of representation with identity; (1) is superstition.

Sure. But it is not necessary that the state of affairs described by (1) ever occur for it to be the basis of an analogy. It is only necessary that one understand what it would be for it to occur. Then one can draw the analogy and add, "now, you see, in the analogically extended case, things really are that way"—as in, "now, you see, the molecules of an ideal gas really do collide with perfect elasticity, even though the billiard balls in my explanatory model do not." It need not even be possible for the analogy-base phenomenon to occur. Billiard balls cannot be perfectly elastic. We do not know whether Descartes believed the state of affairs described in (1) occurs, or can occur. But he believed he understood what it would be for it to occur, surely, or he would not have rested everything on the proposition that it is an intelligible phenomenon.[4]

There is one more step of free-wheeling metaphysics to be taken before we have got the result that a sun-thought, the act, is the same thing as the

sun, its object. The analogical claim (2) seems not to require that the sun-thought be the sun. All it requires is that the sun exists in the thinker's thinking, or, for short, "in the mind," and that this happens because the thinker experiences her sun-thought as the sun and manipulates her sun-thought mentally as the sun. A weaker identification appears to accommodate all of this:

(ID1) the presence in the mind of the sun-thought = the presence in the mind of the sun.

Or, loosely, a thinker having her sun-thought in her thoughts is the same thing as her having the sun in her thoughts. One state of affairs is the other state of affairs. But there is a line of thought in traditional metaphysics that reduces (ID1) to:

(ID2) the sun-thought = the sun.

The plausibility of (ID2) derives from an intuition it is easy to feel. There is, clearly, a difference between (i) an object X, and (ii) the object X existing, but it is a kind of fake difference. It is a difference our grammar twists our arms into making. Really, "in the things" there is no difference. This holds for existing in the mind just as much as it holds for existing in reality. To the extent that it is hard to find a "real" difference between a cat (in your room) and the presence of that very cat (in your room), it is hard to find a difference between the cat (in your mind) and the presence of that very cat (in your mind). More than one philosopher has made a lot rest on this idea, though usually the point has been made specifically about existence-in-reality, rather than about existence in general (including existence in the mind). Hume made the point this way (in the course of an argument for his preferred conception of belief):

> 'Tis also evident, that the idea of existence is nothing different from the idea of any object, and that when after the simple conception of any thing we wou'd conceive it as existent, we in reality make no addition to or alteration on our first idea. Thus when we affirm, that God is existent, we simply form the idea of such a being, as he is represented to us; nor is the existence, which we attribute to him, conceiv'd by a particular idea, which we join to the idea of his other qualities, and can again separate and distinguish from them.[5]

The point here, as in the similar remarks by Kant, concerns "existence in extramental reality." But the force of the point arises from the fact that

extramental reality is a kind of *being*. The underlying thought is that "in reality"—whatever mode of reality you mean to discuss—there is no difference between an object and the presence, the being, of that object. (This metaphysical intuition is wiped away by the usual linguistic reformulation of Hume's, or Kant's, points—the kind of reformulation we considered in Chapter 9.)

As against this intuition, it seems there is *some* sort of difference between an object and the object being present, since the being-present of an object is in the logical category of a state of affairs, whereas the object may be of any logical category whatever. But one has a strong inclination to say "Sure, sure, reason can find a difference—but what difference is there *really?"* It is plausible to think there is nothing here but a disagreement over how seriously we should take the grammatical difference between "the being of the sun in the mind" or "the presence of the sun in the mind" on the one hand, and "the sun (in the mind)" on the other. If one takes it seriously, one is making a "distinction of reason"; if one does not take it seriously, one is declining to distinguish among things according to distinctions of reason. Assume we decline.

Then we have:

the being (in mind) of the sun = the sun (in the mind).

It seems that "the being in mind of the sun" is just another way of saying "the presence in the mind of the sun." So:

the presence in the mind of the sun = the sun (in the mind).

Earlier we reached (ID1):

(ID1) the presence in the mind of the sun-thought = the presence in the mind of the sun.

But now we may infer from this:

the presence in the mind of the sun-thought = the sun (in the mind).

Surely "the presence in mind of the sun-thought" is just another way of saying "the sun-thought." Thus we wind up with:

(ID2) the sun-thought = the sun (in the mind).

This completes the identification of act and object. In the logic of the position, the "category mistake" of identifying an act of awareness, which is a state of affairs, with a particular (an intramental particular, but still a par-

ticular), occurs in the reduction of (ID1) to (ID2), a reduction licensed by rejecting (some) distinctions of reason, distinctions forced upon us by the mere grammar of our language. Instead, we accept certain intuitively compelling identifications that oblige us to fly in the face of grammar so as not to falsify the things themselves.

The founding analogy made the act-object identification seem natural. But it is not obvious from the founding analogy alone how identifying act with object was supposed to solve any traditional philosophical problem. It is not obvious what one might use it for, once one had it available. There is a general answer to the question what it is good for, an answer attractive to rationalists and empiricists alike. And there are specialized answers that speak to the interests of, say, empiricists. We'll look at a specialized answer later. Here is the general answer.

To repeat, the theory of ideas was a theory of the aboutness of thought: some thoughts are awarenesses, graspings, representings, of objects. Some objects one thinks about are far away, some do not exist (in reality). Therefore the theory of ideas owed an account of how a thought can be "of" or "about" an absent object, or a nonexistent object (an object which does not exist in reality). The theory of ideas explained thought-in-absence by denying it. When you think about the sun, the sun exists in your mind, in the manner in which things exist in minds, so that it really isn't absent after all. The sun "exists in your mind" by *being* your sun-thought; the latter, clearly, is in your mind. Intentionality is identity.[6]

Actually, this doctrine is not quite as radical a departure from traditional scholastic doctrine as at first glance it might seem to be. One strand in scholastic theory, the theory Descartes and his successors rejected and replaced with the theory of ideas, also analyzed intentionality as identity: the relation between a species out in nature, and that same species "impressed" in the potential intellect, was identity (this was denied loudly by some proponents of the view, but it is hard to see the ground for these denials).[7] But this traditional scholastic theory was good only for objects that happen to be predicables, things that can be predicatively "had" by other things.

A quick review: the point was to say how the "knowable" or understandable elements of reality are stored in the mind (intellect). What was "knowable" in the first instance was a universal, a form. A form could be in the mind (the potential intellect) by informing it. Forms were not exactly like what we call properties, since our properties are exemplified by "complete" things such as tables or beliefs. Forms are instantiated in "matter,"

which is incomplete precisely in that, taken by itself, it is missing a form. But we will not be far wrong if we think of forms as exemplifiable items, like properties. The form RABBIT can be exemplified in rabbit-matter (guts and stuff) out in the field, and also can be exemplified in intellectual matter in one's mind, permanently at hand to the thinker as a "species impressa." You don't get a furry, smelly rabbit when RABBIT is exemplified in intellectual matter. So the answer to the question "How is the universal *rabbit* in one's mind" has an answer: it is *had* by one's mind, though not in a smelly way.

That never was a good explanation of how a particular, as opposed to a universal, could "exist in" a mind. Particulars cannot be exemplified by things, including minds. The early modern philosophers were bound and determined to allow thoughts to be about particulars, such as oneself or the last light-particle to hit the snowball (it is too big a question why this was true; if the reader disagrees, let me stipulate it). So the theory of ideas had to explain how a particular can be in a mind.[8] The answer—a thing, whether it is a universal or a particular, exists in the mind by being a thought—does not sound like the scholastic doctrine. But consider: the thing, the object, exists in the mind the way a thought does. A thought is a modification of the mind, a way the mind is. A modification of the mind is "had by the mind" much as a form is "had by the intellect" (although the logic of "modifications of substance" is a little different from the logic of forms). Still, one can say that the bottom line is that an object exists in the mind by being had by the mind. The theory of ideas and the traditional (Thomist) scholastic theory posit very similar "mechanisms of attachment" in order to explain how an intramental object is "attached to" the mind. The nontraditional step in the early modern account is just the cross-category identification of the-sun-in-my-mind with my sun-thought, the rejection of a distinction of reason, the rejection of grammar as ontologically significant.[9]

The need to explain thought-in-absence was shared by every proponent of the theory of ideas, and was the single most important philosophical job assigned to the act-object identification. But to make the point that inferential charity can be well motivated when we are reading these theorists, it is best to have an example of a philosophical job where the labor is neatly divided between an "act" component of the job and an "object" component. I'll give two examples. Both are more specialized jobs than the job of explaining thought-in-absence; only some of the early modern philosophers wanted these jobs done at all.

First example: two themes were typical of the theory of ideas, though not every proponent emphasized both of them. One theme had to do with the structure of ideas, while the other had to do with the "worldly fit" of ideas. It was hard to bring these two aspects of ideas together; but many important arguments needed to bring them together—somehow. I'll work with Locke's version.

Theme (1). Locke classified ideas according to structural simplicity. "Complex" ideas have simpler ideas as components, resolving finally into "simple" ideas which have no components other than themselves. Locke tried to explain the grasp of concepts (as opposed to the aboutness of concepts) with the help of this machinery, and laid down rules for the accurate, nonmisleading use of language with the classification of ideas as (relatively) simple or (relatively) complex as a backdrop.

Theme (2). Locke held that physical-object ideas do not resemble physical objects in all respects, although they resemble them in some respects. For example, ideas do not resemble physical objects in respect of color, though ideas do, or may, resemble physical objects in respect of shape. This was a characteristic early modern opinion, and usually led, as it did for Locke, to a partial skepticism about the possible scope of science.[10]

Both kinds of idea-feature, the "structural" feature and the "worldly fit" feature, were ascribed to the same class of mental items—ideas. But the structural feature usually is best thought of as a property of mental states or episodes, whereas the worldly fit feature usually is best thought of as a property of objects (such as the sun, or a circle) within the mind. Only if ideas somehow "are" both states of mind and objects in the mind can they carry all this weight. The argument goes as follows:

If one takes ideas to be sensory awarenesses—episodes—a simple/complex distinction makes very good sense. Episodes usually have proper parts, also episodes. A running play in a football game has the tailback's activity as a proper part, the pulling guard's activity as another proper part, and so on. Standing states have parts in a similar sense. It is important here to distinguish between the part-events of an event and the nonevent components of the event: the tailback is a nonevent component of the running play; his run is a part-event of the running play.

Now suppose you have a visual image of a red tomato. Your red-tomato awareness is a state of mind, or a mental episode (it does not matter here which of these we think is the more apt description). This mental episode can be called "complex" relative to its (proper) part-events, in that it contains them and they do not contain it. For instance, your red-tomato-

awareness has a red-awareness and a bulgy-round-thing-awareness as proper parts, among others. But the red-awareness you have as a part of your red-tomato-awareness does not seem to have any proper parts in the same sense. Or, if "red" is still too determinable for that to ring true, consider your red1-awareness, where red1 is exactly the determinant of red characterizing your mental image of the tomato. It is possible to fuss about this claim, and critics did. But it has an intuitive plausibility. Obviously, in order for ideas to be "simple" and "complex" in this sense, they must be episodes, acts.

Locke accepted (2) because his model of scientific knowledge required that people be able to accurately represent the features which, when they occur as features of microentities, are operative in physical change. Like other mechanical philosophers, Locke took these operative features to be the "mechanical" features, the features a mechanic (what we would call a mechanical engineer) would need to know about in order to *build* a rock or a pool of blood. For the mechanical philosophers the paradigm of an "accurate representation" of the world was a mechanical drawing, or perhaps an architect's elevation.[11] This meant that in order to have a shot at acquiring scientific knowledge, a person had to possess technical-drawing-like mental representations—ideas that represented *some* of the properties of the world accurately, in the sense that by applying a known "scale" and some sort of perspective-correction, one could infer the worldly properties from the properties of pieces of the idea, and infer worldly relations from relations among pieces of the idea. The properties that these ideas had to get right were shape, spatial relation, and the other properties of mechanical, including dynamic, importance.[12]

So Locke insisted that ideas can resemble physical things in respect of the features important in mechanics. It would seem that an idea must be an object-in-the-mind (a circle-in-the-mind, for example) for it to *share* "mechanically relevant" properties, such as shape, with an extramental physical thing, whether the thing is micro or macro.

Now suppose you have some largish swatch of reasoning from Locke's *Essay,* and it draws upon both aspects of ideas at once. That is, in this swatch of reasoning Locke relies (explicitly or by presupposition) upon both his views on idea-structure and his views on the worldly fit of ideas. It would be hard to find a major argument in his philosophy of science or his philosophy of language that does not fit this bill. Suppose, further, we adopt a semantics for Locke's language of ideas according to which he sometimes used the word "idea" to denote a mental state or episode, and

sometimes used the word "idea" to denote a thing-existing-in-the-mind. This semantics probably will interpret the swatch of reasoning we have in mind as a trivial fallacy of equivocation, since it spans the gap between one sense of "idea" and the other—indeed may leap back and forth many times. Too bad for Locke's philosophy of science and philosophy of language.

To put the point another way, if we interpret Locke as using "idea" ambiguously, we will be obliged to conclude that he *has no account* of the structure of "ideas" in the sense of objects-in-the-mind, and *has no account* of how "ideas" in the sense of episodes of awareness ever accurately depict the physical-thing properties they must accurately depict according to his model of science.

By contrast, if we hold that Locke identified awarenesses with the intramental objects of awareness, if we hold that he identified act and object, it seems we have avoided all this trouble. But then we must accept the duty of explaining what it means to say he identified these two disparate things. My suggestion is that saying "Locke identified act and object" is a way of taking the semantic position: "Let's appraise Locke's arguments for validity by means of a suitably charitable validity criterion, one that does not penalize him for failing to discriminate between act and object in his reasoning."

Obviously, a detailed example is in order. Rather than examine in detail some Lockean argument that invokes both the "structural" and the "worldly fit" aspects of ideas, let's switch to a different example, but one that makes the same point.

19

Making Category Mistakes and Loving It

The argument we are about to examine is Locke's defense of his anti-Cartesian contention that no idea is confused, and that this fact actually provides the basis for a type of incorrigible propositional knowledge. I think it is amusing that in order to defend his view (that no idea is confused, and that this has further epistemological implications) Locke gave an argument we can ratify as valid only by taking him to be operating with a strategy of self-induced confusion. We have met the argument before, though not in full dress.[1] It is ultimately an empiricist argument, and therefore of special importance only to some early moderns. Here is the first part of a two-part swatch of reasoning:

> *Part (1):*
>
> Every one that has any Knowledge at all, has, as the Foundation of it, various and distinct *Ideas:* And it is the first act of the Mind, (without which, it can never be capable of any Knowledge,) to know every one of its *Ideas* by it self, and distinguish it from others. Every one finds in himself, that he knows the *Ideas* he has; That he knows also, when any one is in his Understanding, and what it is; And that when more than one are there, he knows them distinctly and unconfusedly one from another. Which always being so, (it being impossible but that he should perceive what he perceives,) he can never be in doubt when any *Idea* is in his Mind, that it is there, and is that *Idea* it is; and that two distinct *Ideas,* when they are in his Mind, are there, and are not one and the same *Idea*.[2]

The parenthetical remark that "it [is] impossible but that [one] should perceive what [one] perceives" seems to be a precis of this claim made earlier in the *Essay:*

> let any *Idea* be as it will, it can be no other but such as the Mind perceives it to be; and that very perception, sufficiently distinguishes it from all other *Ideas,* which cannot be other, i.e., different, without being perceived to be so.[3]

As I remarked in Chapter 2, this is Locke's form of the incorrigibility of the mental. Ideas are "given as unconfused." This is not a thesis about propositional knowledge; it is a thesis about the mind's capacity to escape confusing one thing with another at the level of ideas. One cannot claim that if rock R1 appears different from rock R2, then the rocks are different; nor can one claim that if rock R1 and rock R2 appear identical (that is, judging from appearances "R1" and "R2" are different terms for the same rock), then the rocks are identical. But the analogous claims *can* be made with respect to ideas (says Locke). Evidently this is supposed to follow from, or be equivalent to, the fact that if red appears different from blue to you, then your idea of red is different from your idea of blue, and vice versa. That makes sense if an idea *is just an awareness.* Being "differently aware" would then imply having different ideas, and having different ideas would imply being differently aware.

Straightaway Locke points to the fact that no idea is confused "in itself" to explain how people come by a vast body of absolutely certain propositional knowledge. Call this Part (2) of the overall swatch of reasoning:

> *Part (2):*
>
> a Man sees the same *Idea* to be the same *Idea,* and infallibly perceives two different *Ideas* to be different *Ideas.* For when a man has in his Understanding, the *Ideas* of *one* and of *two,* the *Idea* of *Yellow* and the *Idea* of *Blue,* he cannot but certainly know, that the *Idea* of One is the *Idea* of One, and not the *Idea* of Two; and that the *Idea* of Yellow is the *Idea* of Yellow, and not the *Idea* of Blue. For a man cannot confound the *Ideas* in his Mind, which he has distinct.[4]

As a consequence, one can know, and know with certainty, such identities and differences as "whatsoever is white is white," "a man is not a horse," "red is not blew."[5]

As far as I know, Locke's full argument—Part (1) followed by Part (2)—is the only argument ever given by a British empiricist to rebut Descartes' conception of common sense as confusing real, physical features of the world with features that occur only as the content of acts of mind; for example, "mechanical" coldness and the cold way of feeling that a snow-

ball presents to the mind. Descartes was happy with the conclusion that immediately follows: common sense *stands to learn what its concepts are* as science clears away the confusions. But it was foundational for British empiricism that common sense has its own autonomous set of concepts, assembled from primitives given in sense-experience, and it was foundational for British empiricism that a person equipped with common sense is in a position to know a priori, by scientifically uninformed reflection upon those concepts, that "red is not blew" and the like. Therefore it was anathema that common sense might be unable to know what its concepts are—in the sense in which Fred does not know what thing is before his eyes when he stares at "Charley"—and therefore be unable to know whether red is or is not blue.[6]

But Part (1) of Locke's reasoning looks to be valid just in case ideas are awarenesses, mental episodes, or states. If that is how we continue to understand ideas in Part (2), how exactly is it supposed to follow from the premise "My awareness of red is not my awareness of blue" that red is not blue? Or, to turn it around, how is it supposed to follow from "My awareness of color X *is* my awareness of color Y" (that is, I have exactly one awareness-type for both X and Y) that X *is* Y? For a quick counterexample to the first inference, imagine that you have a piece of oak painted red that you use as a sample of that shade of red, except when you get bored using it and instead use your piece of aluminum painted the same shade of red. The "representative items" are different, "materially" different, but the property represented is the same. Alternatively, suppose you use your red-painted piece of aluminum as a sample of red and also as a sample of chromatic color, as the mood strikes you. One representative item, two objects. So it is in general with representation, with "ofness."

Part (2) of Locke's reasoning draws conclusions about the objects of awareness, so at this stage of his reasoning Locke must be thinking of "one's idea of red" as red-existing-in-one's-mind. But Part (1), the argument that one cannot confuse ideas, seemed to require that ideas be awarenesses—states of mind or mental episodes—so that one's idea of red is one's being-aware-of-red. Locke is convicted of equivocation, and his reasoning taken as a whole is trivially invalid.

Or rather, that is what we are bound to say if we accept the semantic characterization presupposed by my gloss. I was interpreting Locke as using "Idea of X" to *refer to* something, and to refer singly, not "multiply." And I was assuming that argument validity has a classical meaning of truth preservation. *Given* that semantic characterization, in order for part (1) to

stand a chance of being a valid argument, "Idea of X" had to refer to one's awareness of X. And in order for Part (2) to stand a chance of being valid, "Idea of X" had to refer to X, or X-in-one's-mind. Then Parts (1) and (2) together comprise a trivial fallacy of equivocation.

We have another semantic option. We can take a position of inferential charity, just as we did toward Fred. We can take the position that Locke should have his reasoning appraised for validity by means of a suitably forgiving criterion of validity.

And what recommends inferential charity? Why should we not just dismiss Locke's argument as a fallacy of equivocation, however influential it may have been on the development of empiricism? The answer is that Locke had an interesting philosophical position, but it can only be appreciated if one can appreciate positions that embody category mistakes. The philosophical position goes like this:

First, assume that when you are aware of red, this awareness of red is also an awareness of itself. It is epistemically reflexive. This does not strike us as especially plausible. It does not even sound like the right way to put the point—we would rather say that every awareness of red is accompanied by a "second-order" awareness that the "first-order" awareness of red is occurring. And then we would deny it. (Most of us would, anyhow.) No matter. Grant me, please, that every early modern empiricist thought it too obvious for words that awarenesses are reflexive knowings.

As soon as you make this assumption, you have at hand a limited but useful type of "incorrigible" knowing—whenever a person experiences an awareness of red, she knows she does. Awarenesses of red never go undetected; when one occurs, it is noted and identified for what it is. What I mean by "identified" is this: you never will misidentify an awareness of red as something else (some other awareness). You never will confuse an awareness of red with something else.

Now make another assumption: identify your awareness of red with red, the property, existing in your mind. (That is, make this special case of the act-object identification.) Plugging this identity into the conclusion of the last paragraph, we get the result: you never will confuse red, existing in your mind, with something else. In this way your reflexive knowledge of one of your "inner experiences"—an awareness of red—is transmuted magically into identificatory knowledge of a property. Moreover, it is identificatory knowledge of the property that is guaranteed to be confusion-proof—you know which property it is with no possibility of confusing it with another.

The leap from knowing your states of mind (your awarenesses) to knowing properties, even properties-existing-in-your-mind, is (to fall back on a slogan) the leap from subjectivity to objectivity. Red-existing-in-your-mind is the kind of thing the world is made of, whereas states of mind are the kind of thing minds are made of. If you are John Locke, you will go cautiously here, since although red-existing-in-your-mind is the "kind of thing" the world is made of in the sense that it is (i) a property and (ii) not a property of minds, you (Locke) believe that nothing in the world ever actually has that property. If you are Locke, you will believe that more steps of construction are required to obtain a property that things in the world do have (steps of construction involving the concept "power" and perhaps other causal concepts). But the main point remains: you can have unconfused knowledge of a nonmental property, an object of thought rather than a thought-state.

Moreover, this knowledge is prescientific. You need not await the teachings of science to learn *which properties, if any, you are thinking of*. Descartes therefore is wrong. In principle, one ought to be able to cobble together a concept of "red" applicable to sunsets, and to reagents changing color in a flask, and applicable to these things by a person who does not yet know what scientific theories to accept. Empiricism, therefore, is possible.[7]

All the rest of the familiar skeptical worries for empiricism come *after* this point: worries about how to construct intersubjective concepts from these materials, or worries about how to find out whether these intersubjective concepts apply to anything "outside all minds." But without this first move, taking us from awareness of an awareness to awareness of an object, there would be no reason to discuss these familiar skeptical worries. There wouldn't even be anything on the *inside* of the veil of perception. More precisely, there wouldn't be any *things* on the inside; things of which we can ask "do they also exist without the mind," or "are they just like things existing without the mind," or "do they exist in the minds of many people at once." On the inside there would be only the rustling of the minds themselves.

This is one of the more elegant epistemological theories. It needs scrutiny, and today most people would agree it needs to be rejected. But it is cogent, it sticks together logically. I have played it out at such length because the only proof that it sticks together logically is to run through the reasoning. It is worth trying to appreciate the main reasons driving this theory. In the end, one may reject these reasons, but the rejection cannot be on the grounds that as reasons they are *trivially* fallacious.

But to avoid trivializing the reasoning of the theory—to avoid having to dismiss it as patently fallacious (a patchwork of fallacies of equivocation)—we must apply an inferentially charitable semantics. It may be that the type of epistemic semantics described in Part Five of this book can be further developed so as to fit the bill. Further development certainly is required. Let's take a simple example of a chain of inferences in the spirit of Locke's reasoning, but written out so as to expose more structure:

(A) If it is necessary that to be aware of red is not the same thing as to be aware of blue, then it is necessary that every person perceives that the idea of red is not the idea of blue.
(B) It is necessary that to be aware of red is not the same thing as to be aware of blue.
So,
(C) It is necessary that every person perceives that the idea of red is not the idea of blue.
Furthermore,
(D) If it is necessary that every person perceives that the idea of red is not the idea of blue, then every person knows that the idea of red is not the idea of blue.
So, from (C) and (D),
(E) Every person knows that the idea of red is not the idea of blue.
So,
(F) Every person knows that red (existing in her mind) is not blue (existing in her mind).

The general semantic strategy would be to posit a Sam and a Sal who have the job of recommending for or against sentences in Locke's "idea" language. Sam interprets "idea of X" as meaning "awareness of X." Sal interprets "idea of X" as meaning "X, as it exists in the thinker's mind." Sam and Sal discharge their responsibilities here just as they did for Fred. In the end, we milk a semantic value out of their joint report on each atomic sentence. Maybe the sentence "The idea of red is the idea of blue" gets the value **N.** One hopes so, or we really need some new authorities. Finally, we sit down and apply some recursive rules that dictate what semantic values should be assigned to compound sentences once all the atomic sentences are assigned values.

But it is obvious that rules described in Chapter 13, sufficient when compound sentences are only formed by mixing in the three connectives "not," "and," and "or," do not suffice to handle *any* of the sentences

in the chain of reasoning (A) through (F). Those sentences (i) contain quantifiers; (ii) contain modal terms; (iii) contain "epistemic" terms such as "perceive" and "know." The phrases we care most about philosophically, that we are most concerned to interpret properly in our semantics, are those containing the word "Idea," and such phrases frequently occur in the scope of modal terms or epistemic terms or quantifiers—or all three (as in "it is necessary that every person perceives that the idea of red is not the idea of blue"). Thus we do not know how to apply the Belnapian validity criterion of Chapter 13 to the subarguments of the chain of reasoning (A) through (F).

Of course this really does not put us in a worse position than we would be in if we wanted to make Locke out to be equivocating, and we were determined to give formal details of a semantics based on the classical concepts Reference/Truth/Truth-of in order to make the charge stick. There is no general agreement among intensional logicians how one should semantically characterize a language which permits nesting of modal operators, epistemic operators, and quantifiers in arbitrary permutations, even if one stays entirely within the family of semantic concepts "denotes," "true," "true-of," "true-at (a world, an epistemic alternative, and so on)." That would not prevent us from drawing a great many fairly confident philosophical conclusions about the semantics of the chain of reasoning (A) through (F). For example, we can see that it is very likely that in order for (A) to be *true,* the term "the idea of red" must *denote* one's awareness of red, and the term "the idea of blue" must *denote* one's awareness of blue. And we can see that it is very likely that in order for (F) to *follow validly from* (E), the terms "the idea of red" and "the idea of blue" in (E) must *denote,* respectively, red-in-one's-mind and blue-in-one's-mind. So we can see that it is very likely that the argument "(A), (B), (D), therefore (F)" either is *invalid* or has a *false* premise—the heart of the charge of equivocation one is led to by this uncharitable semantics.

All of these semantic judgments are made in the absence of a worked-out semantics for the language in which the chain of reasoning (A) through (F) is formulated. They are educated guesses about the semantic properties of various bits and pieces of the chain of reasoning, or rather, they are educated guesses about the semantic properties a worked-out Reference-Truth-based semantics would ascribe.

We cannot be quite so confident in our educated guesses about the semantic properties a worked-out Belnapian semantics would ascribe to bits and pieces of the chain of reasoning (A) through (F). It is not so hard

to imagine quantifier semantics done substitutionally, so that a universal quantification behaves semantically "like an infinite conjunction of substitution instances." Except for the unfamiliar semantic rule for conjunction, this assimilates quantifier semantics to a familiar model and makes educated guesses possible. But standard semantic treatments of epistemic language, language employing such words as "know" and "perceive," use the Reference/Truth/Truth-of cluster of semantic concepts, as do standard semantic treatments of modal language. It is less intuitively clear how to bend these standard treatments around so as to mesh with Belnapian epistemic semantics.

So let's proceed this way: let's assume there is some way to extend the Belnapian epistemic semantics of Chapter 13 to apply to the rich philosopher's language used by Locke, some of which ordinarily would be called "nonextensional" language.[8] This may be a false assumption. If so, we must finally retract our claim to understand Locke's philosophy—or else find another suitably charitable criterion of validity for his reasoning. But assume an elaborated version of the Belnapian semantics will work. This means we are entitled to make only rather general educated guesses about the semantic properties of Locke's statements and arguments. But we will be safe if we limit ourselves to philosophical claims that are implied by "the very idea of" Belnapian epistemic semantics. For example, however the details are worked out, the semantics will oblige us to think of Locke's statements, his beliefs, as **Y, N, Y&N,** or **?,** rather than as True or False. This is enough information to yield some philosophical payoff.

It goes without saying that the semantic palette **Y, N, Y&N,** and **?** is unfamiliar. In particular, the value **Y&N** is unfamiliar. The other three values can play rather more familiar roles in a philosophical critique. One might applaud Locke for somehow managing to hold many **Y** beliefs, despite his self-induced confusion; because it is epistemically praiseworthy to believe the **Y,** just as it is epistemically praiseworthy to believe the True. One might chide Locke for believing too many **N** things on a particular topic, on the grounds that believing the **N** is bad and cannot be fully explained as a consequence of confusion. The value **Y** is truelike and the value **N** falselike, with respect to the role they play in epistemic critique (though of course not with respect to ontological significance).

Even the value **?** can play a familiar role in epistemic critique. We may prefer to teach our beginning students two-valued logic, but when we switch classrooms and teach them about dead people we make ample use of a neither-true-nor-false value. One could think of this as the intermedi-

ate value of a Weak Kleene Scheme, with truth and falsehood the other two values.[9] When a dead philosopher never anticipated a certain later development, and one does not know what to make of some of the philosopher's views in light of the later development, it is quite natural to declare the question off limits, and decline to ascribe either truth or falsehood.[10] In a three-valued semantics, one can ascribe the third, neither-true-nor-false, value to such sentences and thereby learn from the semantics how they "fit into" implications. In a Belnapian epistemic semantics, the value ? can play much the same role in epistemic critique as the neither-true-nor-false value of a three-valued logic (though here again the ontological significance is very different).

But the value **Y&N** has no parallel in the Reference/Truth/Truth-of cluster of semantic concepts. A formal contradiction, for example, is plain false, not "both true and false." You can't be "both true and false." Suppose Locke states some philosophical belief, using the language of ideas, and the belief is **Y&N.** Then its negation also is **Y&N.** The semantic value **Y&N** does not mark a belief as one that we should accept (and not reject), or reject (and not accept). Nevertheless, it is wiser to accept a **Y&N** belief than it is to accept an **N** belief, although it is wiser to accept a **Y** belief than it is to accept a **Y&N** belief. Locke might hold a large number of **Y&N** beliefs, all of them "not wholly unwise to accept"—just as their negations are not wholly unwise to accept.

This system of beliefs might be supported by subtle and valid arguments, arguments beginning with **Y&N** premises (perhaps with the occasional **Y** premise scattered about). The argument we examined above may be an example, if indeed we can extend the Belnapian semantics to fit the rich modal and epistemic language in which the argument is expressed. A system of philosophical beliefs, all valid consequences of various combinations of (mostly) **Y&N** beliefs, and having, in turn, various other **Y&N** as valid consequences, could be an interesting piece of philosophy. But since marking beliefs as **Y&N** does not give us any information about what we should accept and not reject, it does not entitle us to take a stand one way as opposed to the other. Read the scope right: it does not (entitle us to take-a-stand-yes while obliging us to not-take-a-stand-no); and it does not (entitle us to take-a-stand-no while obliging us to not-take-a-stand-yes). By contrast, marking beliefs true or false *does* (entitle us to take-a-stand-yes while obliging us to not-take-a-stand-no)—if the belief is true; or else it *does* (entitle us to take-a-stand-no while obliging us to not-take-a-stand-yes)—if the belief is false. One should agree without disagreeing with the

true, and one should disagree without agreeing with the false. But our charitable semantics, if we can actually produce the details, has no truck with truth and falsehood.

Usually we expect that valid arguments "to" and "from" a belief will serve us as guides for deciding whether to agree rather than disagree, or disagree rather than agree, with the belief itself. Valid arguments connect a given belief to others, so that eventually, when we put together a big enough quilt of argument, we will see how things we already agree with or disagree with require, rationally, that we agree with (and not disagree with) or disagree with (and not agree with) the given belief. We expect valid arguments to pile up until we learn more and more about where to take a stand.

But this may not happen if we are applying a Belnapian semantics to some sector of Locke's (or whoever's) philosophy where act-object identification is prominent. We choose the charitable semantics in order to appraise fairly the philosopher's stated reasons for various beliefs, the stated consequences of those beliefs, and in order to make reasonable suggestions concerning unnoticed consequences of the beliefs. Then valid consequence upon valid consequence pile up, linking the beliefs into a "system" of considerable interest, maybe even a system in which no proposition is wholly unworthy of acceptance. But none of this gets us any nearer to settling what to affirm rather than deny (or vice versa). Things might not work out that way, but then again they might.

When we say "I understand so-and-so's views," or "I know what so-and-so thinks," we presuppose that there is an intact linkage between (a) learning more and more about the valid consequence relations among sentences in the subject's language; and (b) learning more and more about which sentences of the subject's language one should affirm (only) or deny (only). Part of understanding is appreciating reasons; part of understanding is knowing whether to agree, knowing how to choose between agreement and disagreement. When this linkage is severed, as it is, or may be, when we apply our charitable semantics to Locke, we are left with the sense that we do not understand his views; that we do not know what he thinks. This accounts for the felt "unintelligibility" of his views. But it is compatible with this source of "unintelligibility" that we can accurately appraise his arguments for validity, tell what follows from what. This untangles the "metaphilosophical puzzle" I posed early in Chapter 18: how can beliefs that are somehow unintelligible fit together into perfectly good reasons for holding beliefs—maybe even deep and exciting reasons? I sus-

pect the unintelligibility of category mistakes ought always to be explained along these lines, but that is another story.

Notice that this untangling of the metaphilosophical puzzle makes the "unintelligibility" of each of Locke's beliefs taken singly a reflection of a global feature of his system of belief—a pulling apart of two epistemic features that (typically) characterize one's grasp of a system of beliefs—improving grasp of entailments and improving grasp of what to accept and reject. That is why we find **Y** sentences as "unintelligible" as **Y&N** sentences. It is worth having a concrete example in mind, so assume Susan is picturing four green discs with a fifth in the center of the array. Consider these three sentences, and in each of them attach to the term "idea" the meaning John Locke attached to it:

(1) Susan's idea of an array of five green discs has Susan's idea of green as a part.
(2) Susan's idea of an array of five green discs resembles that array of plastic discs on the wall in respect of geometric/topological property P (the array on the wall consists of a square of four real discs with a fifth centered, so any number of geometric/topological properties will do the job).
(3) Susan's idea of an array of five green discs is an abstraction from particular experience.

Presumably Locke would endorse all three of these sentences. (1) sounds pretty good if you think of the expression "Susan's idea of an array of five green discs" as referring to an awareness-episode, an act, and if you think of "Susan's idea of green" as referring to another, simpler awareness (perhaps one of a type that can only occur as a part of more complex awarenesses). Sam will like that sentence. But Sal won't. Like every baffled undergraduate who can't figure out how a visual image of, say, an array of discs can have "an image of green" as a "simple component," Sal will reject (1).[11]

The vote will go the other way around on (2). The sentence makes sense if the expression "Susan's idea of an array of five green discs" refers to an object, a disc array in Susan's image (analogous to the boat in the painting). But it makes no sense if "Susan's idea of an array of five green discs" refers to a sensory episode.

Sentence (3) has quite a good "act" interpretation—Susan has learned to be in this type of sensory state by being in instances of it, and these instances shared a particular feature. It also has quite a good "object" inter-

pretation—Susan's array-of-green-discs image is a simplified copy of various particular perceptual images. Both Sam and Sal will like (3). (Of course they might both reject abstraction, however interpreted, as an idea-forming mechanism. But let's ignore that possibility so that we can have fairly simple examples.) The upshot is that (1) and (2) both have the semantic value **Y&N,** and (3) has the semantic value **Y.**

Intuitively, (1), (2), *and* (3) share equally in the "unintelligibility" induced by the act-object confusion; all of them suffer from the category mistake. When a sentence is **Y,** when it is plausible on both interpretations of "idea of X," one might have expected the bite of the confusion to be weaker than when a sentence is **Y&N,** plausible on one interpretation but not on the other. Intuitively, this is not what happens. We need an explanation.

And we have one. (1), (2), and (3) are equally members of Locke's system of belief, a system for which two epistemic system-features pull apart. Our ability to learn more and more about the valid consequence relations among sentences in Locke's philosophical language pulls apart from our ability to learn more and more about which sentences of Locke's philosophical language one should affirm (only) or deny (only). These epistemic features need to be present together for us to comfortably think of ourselves as understanding Locke's beliefs. And the pulling apart of the two features is a global phenomenon; all sentences of Locke's philosophical language are affected (or rather, all sentences of the fragment of his language to which we choose to apply our Belnapian semantics), so it does not matter which semantic value a given sentence has: it is just as "unintelligible," and "unintelligible" in just the same sense, as every other sentence.[12]

It is possible to draw a moral. If one wishes above all else to *evaluate John Locke's beliefs* for truth or falsehood, in order to decide which of his beliefs ought to be reaffirmed and which of his beliefs ought to be rejected, there is no choice but to apply a semantics that marshals the Reference/Truth/Truth-of cluster of semantic concepts. The issue is precisely when one of Locke's beliefs is true-of-reality and when not. The likeliest candidate for an appropriate semantics would seem to be an "ambiguity" semantics, whereby one interprets Locke as using the term "idea" now to refer to acts, now to refer to objects. This semantic choice obliges one to make uncharitable assessments of his *reasons.*

If, on the other hand, one wishes to *appreciate Locke's reasons;* if one wishes to be able to evaluate his arguments for validity in a way that does

not dismiss as trivially invalid most good-sized swatches of his reasoning, then one ought to choose a Belnapian semantics. This obliges one to abandon the idea of assessing Locke's beliefs for truth-of-reality. There is a simple trade-off: appreciate Locke's reasons or evaluate his beliefs for acceptability as true or false.

This moral is general. Nothing special about Locke. Historians of philosophy usually care a great deal about the reasons given by smart dead people for their views, and care rather less about which of those views to accept, unmodified, as their own. For such historians the correct semantic choice is clear. Of course not all past philosophy depends for its rational coherence on cross-category identification; when no such issue arises, might as well be classical, or at worst three-valued.

So much for how to *read* Locke (or whoever). *Was Locke himself* well advised to produce arguments that can be certified as valid only at the cost of certifying their component beliefs as true (or damning their component beliefs as false, or noting that by contrast with these possibilities those component beliefs are without truth-value)? Are *you* well advised to do it? Is it a reasonable philosophical strategy? We know some of the benefits: you get to work with "merged," cross-categorial objects if you cannot think of a way to avoid them, or cannot *yet* think of a way to avoid them. And even while framed in terms of merged, cross-categorial objects, your thought can be entirely *reasonable*. You can wire together an inferentially connected system of beliefs for which the inferential links are valid; things follow from things. But we know some of the costs, too. You cannot classify the things you say as true or false, and therefore you cannot classify them as worthy of your belief because true, or unworthy of your belief because false. How should you add that up? It seems to me the question simply must be answered, though not here.

Notes

1. Thinking One Thing Is Another

1. While trying to find a fruitful way to look at the phenomenon of confusion, I found the following especially helpful: Gareth Evans, *The Varieties of Reference* (New York: Oxford University Press, 1982), especially ch. 4; Margaret Wilson, "Descartes on the Representationality of Sensation," originally in *Central Themes in Early Modern Philosophy* (Jonathan Bennett Festschrift), ed. M. Kulstad and J. Covel (Indianapolis: Hackett, 1990), pp. 1–22, and reprinted as ch. 5 of M. Wilson's *Ideas and Mechanism, Essays on Early Modern Philosophy* (Princeton: Princeton University Press, 1999), pp. 69–83; Hartry Field, "Theory Change and the Indeterminacy of Reference," *Journal of Philosophy 70* (1973), pp. 462–81; Bas van Fraassen, "Singular Terms, Truth-Value Gaps and Free Logic," *Journal of Philosophy 63* (1966), pp. 481–495, and also his "Presupposition, Implication, and Self-Reference," *Journal of Philosophy 65* (1968), pp. 136–152; Saul Kripke, "Speaker's Reference and Semantic Reference," *Midwest Studies in Philosophy II* (1977), pp. 255–276, reprinted in *Contemporary Perspectives in the Philosophy of Language,* ed. Peter A. French, Theodore E. Uehling, and Howard K. Wettstein (Minneapolis: University of Minnesota Press, 1979), pp. 6–27; Daniel Dennett, "Beyond Belief", in *Thought and Object: Essays on Intentionality,* ed. Andrew Woodfield (New York: Oxford University Press, 1982), pp. 1–95; Anil Gupta, "Meaning and Misconceptions," in *Language, Logic, and Concepts: Essays in Memory of John Macnamara,* ed. Ray Jackendoff, Paul Bloom, and Karen Wynn (Cambridge, Mass.: MIT Press, 1999), pp. 15–41.

   I had not given a second thought to the phenomenon of ontological confusion until Barbara Hill pointed out its philosophical importance, in conversation, sometime in the mid-to-late seventies.
2. This idea is discussed in some detail in Chapter 6 and then again in Chapter 7.
3. Someone might prefer to use a different idiom here, and say you have a delusion instead of saying you have a myth. Either idiom has the drawback of assimilating your state to much less fleeting, and, in different senses of "serious," more serious mental states. I believe the delusions typical of some psychotic

states should be seen mainly as ontological confusions, rather than as systems of false belief. Because—for reasons that eventually will come clear—I think confused language is not truth-valuable at all, the distinction between an ontologically confused system of thought and a mostly false system of thought has considerable content. I will not discuss the matter of delusion any further in this book, though.

4. In general, a semantics in this sense provides very little information about the meaning of expressions in the language. For example, if one formulates a validity definition in terms of the concept of truth-in-a-model, one cannot, in general, extract an account of the "truth conditions" of sentences, because to do that one needs to describe the "actual" model. As Gupta and Belnap point out, if the language includes, say, the language of physics, one would need to give the true physics. See Anil Gupta and Nuel Belnap, *The Revision Theory of Truth* (Cambridge, Mass.: MIT Press, 1993), pp. 22–25. Later I will propose a certain semantics for a smallish but interesting class of inferences made by confused people. It is by no means a theory of meaning for the language used by these people. If the semantics I propose has anything to do with meaning, it has to do only with the meaning of "and," "or," and "not."
5. Actually, the third value of the Weak Kleene Scheme is just "neither true not false"; one can decide to count sentences as "neither true nor false" exactly when they are meaningless, or one can decide to do so on other criteria. A good discussion of the Weak Kleene Scheme, and of the "meaningless" interpretation of its third value, is in Gupta and Belnap, *The Revision Theory,* ch. 2.
6. We will confront a similar set of issues in connection with confusion. But for reasons that will emerge, the Weak Kleene Scheme will not prove to be the semantics of choice. It will not suffice to augment the classical cluster of semantic values (truth, falsehood, and related properties such as "truth-of" and reference). We will be obliged to move to an entirely different type of semantic value, with a commensurate alteration in our conception of validity.

2. A Little History

1. This little story is borrowed, much altered, from Paul Churchland, *Scientific Realism and the Plasticity of Mind* (Cambridge: Cambridge University Press, 1979), esp. section 3.
2. When an entire civilization lacks the ability to tell X from Y, we have what Hartry Field has called "indeterminacy." Example (due to Mark Wilson): an early seventeenth-century natural philosopher says something about the "weight" of a certain rock. Nobody at that time knows how to distinguish between mass and impressed gravitational force, so when the philosopher assesses the "weight" of his rock, he uses a blend of what we would regard as tests for mass and what we would regard as tests for impressed gravitational force. Some of what he believes concerning weight is good for IGF but not so good for mass, and vice versa. He confuses mass with impressed gravitational

force—except that, unlike you when you confuse the two cars, such modest steps as paying more attention to his surroundings will avail him not at all. And unlike Marvin, he cannot go to school for a few years and emerge unconfused. If we agree to classify the philosopher's problem as confusion, then many cases of confusion cannot be corrected without a shift of conceptual scheme, and therefore may affect every member of a civilization, for many lifetimes. For an assessment of such cases from a "supervaluation" perspective, cf. Hartry Field, "Theory Change."

3. For the contrary view, well-argued indeed, see Cecilia T. Wee, "Material Falsity in Descartes' Meditations," Ph.D. diss., University of Pittsburgh, 1994.
4. What he meant, presumably, was that a confused idea is the material cause of the ensuing false judgment, which has somewhat richer implications than does the non-Aristotelian formula I put in his mouth.
5. Arnauld wrote this in the "fourth set of objections" to Descartes' *Meditations;* the translation is by Margaret Wilson in p. 72 of Wilson: *Ideas and Mechanism*. There are several generally available translations of all six sets of objections to the *Meditations,* with Descartes' replies. One widely respected translation is in the second volume of *The Philosophical Writings of Descartes,* translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge: Cambridge University Press, 1984), pp. 63–383 (the passage quoted above occurs on p. 145 of vol. II). Wilson's translation of this passage agrees in substance with the translation by Cottingham et al., but the reader who has not been over this ground before may be dismayed to learn that the various "standard" translations often disagree on the proper rendering of many epistemologically and metaphysically crucial passages.
6. In Descartes' time the term-of-art "idea" was used in several ways, and sometimes—confusedly—in two or more ways at once. In Part Seven, and especially Chapter 19, I explore the seventeenth-century philosophical strategy of wittingly "confusing" act with object—confusing a representing state of mind with the thing represented. But it may strike some readers as bizarre that Descartes and others sometimes called ordinary, concrete entities in the world—entities such as the sun—"ideas." We are accustomed to using the term "idea" as a rough synonym for "concept." Therefore we think of ideas as more like descriptive paragraphs, as opposed to the scene described by a descriptive paragraph. Seventeenth-century philosophers, both scholastic and nonscholastic, had something like the descriptive-paragraph model of an idea, just as we do. But they also had the scene-described model. One reason for this is that the original Neoplatonic meaning of "idea" as an exemplar in the divine mind had been loosened up and generalized, but not yet abandoned entirely. Roger Ariew and Marjorie Grene remark that a popular example among scholastic writers (including scholastic writers Descartes probably read) was Seneca's story about an artist painting Virgil. The man, Virgil, is an "idea," because he functions as an exemplar for the painter (an exemplar of Virgilhood, or something like that, I suppose). Descartes happily explained himself in the first set

of replies by noting that his idea of the sun is the sun itself, though existing in the intellect the way things normally exist in the intellect. This "way of existing" talk was a slick surrogate for more complicated talk about an "instrumental" role the sun is playing in one's mental life (see Chapter 18), just as Virgil is an idea because he plays a rather narrowly defined instrumental role in the mental life (the creative activity) of the artist. For the Seneca story see Roger Ariew, *Descartes and the Last Scholastics* (Ithaca: Cornell University Press, 1999), pp. 66–67. For Descartes' "sun" comment, see Cottingham et al., *Philosophical Writings of Descartes,* vol. II, p. 75. A good survey of seventeenth-century scholastic uses of the term "idea" is Ariew, *Descartes and the Last Scholastics,* pp. 58–76.

7. That is, to branch "outward" from ideas to worldly things. There is evidence that he thought the identity relation can branch "inward"—he almost comes right out and says that he has two ideas of the sun, at about the time he was saying his idea of the sun is the sun itself (see the Third Meditation, at p. 27 of Cottingham et al., *Philosophical Writings,* vol. II).
8. In Margaret Wilson, p. 74 of *Ideas and Mechanism.*
9. I am using "real" (property) in our sense, not in the technical sense of early seventeenth-century natural philosophy. In the latter sense, mechanists denied that there are "real" physical properties—properties separable from their subjects at least by the divine will (like the accidents of the bread, which, as everyone knew, are from time to time peeled off and hooked onto the substance of the body of Christ).
10. John Locke: *An Essay Concerning Human Understanding,* ed. with an introduction by Peter H. Nidditch (Oxford: Oxford University Press, 1975), p. 597 (bk. IV, ch. VII, 10).
11. In the continental rationalist tradition, confusion-avoidance strategies remained central to the theory of knowledge. For example, Leibniz had a notion of "confused knowledge." Here is his neat summary in the *Discourse on Metaphysics* (in *G. W. Leibniz: Philosophical Texts,* trans. and ed. R. S. Woolhouse and Richard Francks (Oxford: Oxford University Press, 1998), pp. 76–77): "When I can recognize one thing among others without being able to say what its differences or properties consist in, my knowledge is *confused*" (emphasis in original). This contrasts with "distinct" knowledge: "But when I can explain the evidence I am using, the knowledge is *distinct*" (emphasis in original). Moreover, distinctness in knowledge comes in degrees: "Distinct knowledge has different levels, because the notions which enter into the definition usually require definition themselves, and are known only confusedly. But when everything which enters into a definition or an item of distinct knowledge is known distinctly, right down to the primary notions, I call the knowledge *adequate*. And when my mind simultaneously and distinctly understands all the primary ingredients of a notion, it has *intuitive* knowledge of it. This is very rare; most human knowledge is only confused" (emphasis in original).

To get a feel for what Leibniz is calling "confused knowledge," consider a variation on the parking lot story. You are in the parking lot at night in the company of an acquaintance who has never seen your car. She has a philosophical turn of mind, and so do you. You arrive at your car; you are absolutely sure it is your car and you say so. (You say so since people with philosophical turns of mind often make unsolicited knowledge claims.) Your friend argues against you. She suggests that you may have mistaken another, qualitatively almost identical, car for yours. You pause to debate the matter rather than settling it by sticking your key in the lock. First you argue that the likelihood of a qualitative match between your car and another car in this very parking lot is very small. Your friend takes the point. But, she says, the interesting question is not whether you can be said to "know it is your car" in a loose and popular sense of "know." Of course you can. The interesting question is this: If you undertook to describe your car, there is a limit to how much detail you would be able to include in your description even if you had all the time in the world. One *explicitly* remembers only so much about a thing, even a familiar thing such as one's car. Let the likelihood that there is another car in the lot having all the features of your car you can recall and articulate be p; your confidence level then should be about $1-p$. But it is much higher than that. That's what you mean when you say you are "absolutely sure." You agree with your friend. You grant that you cannot put your finger on any feature, or even any combination of features, that fully justifies your very firm conviction that this is your car. But, you say, one need not be able to articulate one's basis for making an identificatory judgment; there is such a thing as *just recognizing a thing as the thing it is,* and as distinct from other things, with a *rightful* degree of confidence that one cannot underwrite by giving individuating descriptions. There is "just knowing how to tell your car."

And most people would say there is. One can try to explain it: maybe we pick up on features we do not know we are picking up on, maybe we have identification routines we do not know we apply. There are other possibilities. The point is that most people believe it is a human possibility, and indeed one that is frequently exercised.

This is what Leibniz called "confused knowledge." He made a workhorse of the concept. Descartes held that prescientific, commonsensical people confuse *states* of things—physical cold and the cold way a thing feels, for example. Similarly, Leibniz believed people have confused knowledge of states of things. They avoid confusion only by being able to inarticulately recognize a given state of things as the state it is. As knowledge grows more articulate—less a matter of "just telling" what state one is thinking of and more a matter of being able to characterize the state—fewer and fewer of the elements of one's characterizing descriptions are just "told." But, he says, rarely is "just telling" expunged entirely. Usually at least some states of things that figure somewhere in a chain of definition are still only confusedly known.

When confusion stopped being a front-burner problem for epistemologists,

other, related concepts came to seem less interesting. Leibniz's concept of confused knowledge is a leading example. I decided not to make a study of "confused knowledge" in this book, principally because it is not itself a type of confusion, though it is relevant to the epistemology of confusion. Hard questions abound, though, and they deserve more scrutiny than they have been getting. For example: as Leibniz understood the matter, one can almost magically duck confusion by inarticulately recognizing something (typically a state or property of things) as the thing it is. Suppose you have written down a careful definition of X as A and B and C (where X, A, B, and C all are states of things, properties). But all you are able to do is recognize A or B or C as a distinct state among others; you have (so far) no description, no characterization, of them. It isn't as though you have recognized A, for instance, by noticing something characteristic of A, but you just haven't written it down or vocalized it; it isn't even that you have noticed something characteristic of A, but as it happens you have no words for whatever that characteristic is. You have recognized A as the state it is without noticing any of its characteristics at all (except in the trivial sense in which being itself is one of its characteristics). Is your purported definition of X really a definition? How does it give you an *inferential* grasp on X? Suppose the answer is that it does not (this wasn't Leibniz's answer). Then even if you keep on pushing back the frontier of your confused knowledge, how do you ever get to a real *characterization* of anything?

12. The argument of this and the three preceding paragraphs was first suggested to me by Tamara Horowitz.
13. More of this in Part Seven, when we will have enough analytical tools in place to make reasonable sense of it all.
14. Immanuel Kant, *Prolegomena to Any Future Metaphysics,* trans. with an introduction by Lewis White Beck (Indianapolis: Bobbs-Merrill, 1950), pp. 37–38. Presumably the immediate target of Kant's remark is Leibniz's doctrine of confused perception, but Descartes is the target at one remove. Leibniz's conception of confused perception was descended from, and very similar to, Descartes' conception of confused common-sense ideas of perceivable features (e.g., cold). It might appear that since Leibniz had a more or less worked-out conception of confused knowledge—how it comes in degrees and so forth—while Descartes did not, the concept of confusion must have played a very different role in Leibniz's thought from any role it played for Descartes (see n. 17 above). But Descartes did believe that completely logical reasoning was possible in the presence of confusion. He did not attach a label to this human capability, though he could have called it "confused knowledge" if he had wanted to. Rationalists get to call rational capacities "knowledge"; that's one of the things that makes them rationalists.
15. Locke, *Essay,* p. 364 (II, XXIX, 5).
16. Ibid., p. 592 (IV, VII, 4).
17. This last trick is the hard one. Once you admit that there *are* entities besides

your ideas, and Locke clearly thought there were, then it might seem that nothing prevents you confusing one of your ideas with some object that is not an idea, or confusing two (or more) nonideas with one another. A defense of Locke at that juncture probably would go along the lines: if you confuse X with Y, then you have a thought about X and about Y; and if you have a thought about some object, this is the same thing as having that object present to the mind *as an idea*. (Again, see Part Seven).

18. Locke, *Essay,* p. 592–3 (IV, VII, 4).

3. Fred and the Ant Colony

1. The story of Fred, like the parking lot story, resembles several "confusion" stories told by D. Dennett and by G. Evans; I prefer ants to pizza parlors or pennies or steel balls (cuter). All these stories turn on similarities in the visible features of material objects. Now, as in the seventeenth century, the more philosophically interesting cases of confusion involve less concrete objects: the story of Marvin's confusion of specific heat differences with temperature differences is a good toy example. I will use the Fred story to make points that I believe apply much more generally, though at several places in later chapters it will be necessary to use Marvin-like examples, and at several other places it will be necessary to get real. See D. Dennett, "Beyond Belief," and G. Evans, *Varieties of Reference.*
2. See S. Kripke, "Speaker's Reference." In Kripke's discussion the term "semantic" has a broader meaning than I assign it: in the phrase "semantic reference," the adjective "semantic" is intended to connote "as determined by the meaning of the shared, public, language."
3. So far as I know, the point that confusion is not belief was first made by D. Dennett.
4. Saul A. Kripke, *Naming and Necessity* (Cambridge, Mass.: Harvard University Press, 1980).
5. Walter Edelberg got me to say this right, finally.
6. In fact what is traditional is to distinguish *de re* and *de dicto* beliefs (or other attitudes), not *de re* and *de dicto* attributions of belief (or other attitudes). That the distinction ought to be drawn at the level of attributions rather than at the level of states has been made clear by Robert Brandom (see, for instance, *Making It Explicit,* Cambridge, Mass.: Harvard University Press, 1994, esp. ch. 8). Nevertheless, in the following I sometimes will write as though the distinction is at the level of states, for brevity. It is easy to translate, so no harm done (here).
7. D. Dennett gives essentially this last argument.
8. Eventually, I will argue that in certain circumstances we should scotch the concepts of reference and truth in a semantic description of confused language, though for reasons different from those briefly mentioned in Chapter 1 in

connection with option 5. As will become clear later, the "certain circumstances" just alluded to include our present inquiry. When reference and truth are off the table as applicable semantic concepts, there is no sense to the suggestion that when one attributes confusion to Fred one is attributing to him a false belief with respect to Ant A and with respect to Ant B, that the one is identical with the other. Nor is there any sense to the suggestion that one is attributing an "untrue and unfalse" belief, *where the basis for that judgment is that Fred employs a multiply-referring term*. So the theoretical argument of later chapters supplements the piecemeal arguments I have just given.

4. The Semantic Use of Psychological Language

1. A child who was very good at doing integral arithmetic almost certainly would have been told something about fractions by whoever got her to be that good. But it is not a pedagogical necessity.
2. Ken Manders suggested this nice example in conversation.
3. More on this approach to a "semantics of confusion" in Chapter 5.

5. Ambiguity

1. Alan Ross Anderson, Nuel D. Belnap, Jr., and J. Michael Dunn, *Entailment: The Logic of Relevance and Necessity* (Princeton: Princeton University Press, 1992), II, pp. 506–541.
2. For a representative example of a supervaluational validity criterion, see Chapter 7.

6. Humoring

1. See Dorothy L. Grover, Joseph L. Camp, Jr., and Nuel D. Belnap, "A Prosentential Theory of Truth," *Philosophical Studies* 27 (1975) 73–125; reprinted in Dorothy Grover, *A Prosentential Theory of Truth* (Princeton: Princeton University Press, 1992), pp. 70–120. See also the other essays in Grover's book.
2. That is, for any given proposition x, assume there is a sentence Sx which expresses x and moreover is the "standard" way of expressing x. Sx therefore is a function mapping propositions unambiguously into sentences. Presumably the propositions outrun the sentences in cardinality, so this is a heavy-handed idealization. If you find it too heavy-handed, you must complicate argument A# even beyond what I have done, so as to get the effect of a "translevel" move from proposition-information to an affirmed sentence. Or you must redo A# as pertaining to sentences, and rely on a truth-definition, as mentioned below.
3. Since each of these instances is "half" of a Tarski T-sentence.
4. I just used the phrasing "Not really true . . . , . . . not really false" to get the ef-

fect of declining to affirm and also declining to deny. My ear says this phrasing does the job. I gather from responses to an earlier draft that the single word "really" isn't enough for some people's ears; they don't distinguish "not really true" from "not true," and therefore hear "not really true" as a denial. Don't.

5. When we imagined you confusing the two cars in the parking lot, Option (3) was attractive—a supervaluational policy for using the words "true" and "false" in connection with your own (earlier) parking-lot thoughts and statements. Perhaps that is because I all but demanded that you try to "identify with" the (imaginary) confused you. How is it possible to identify, to imagine being in that situation? As a guess, the best one can do is adopt a humoring posture toward this earlier incarnation of oneself. If that is right, it would explain why a supervaluational policy for making "true" and "false" judgments seems as natural as it does, and it would explain how one can succeed in this "identification" without feeling the least rational compulsion to regard oneself, now, as confused.
6. It is an interesting question whether the concept of language "fitting the world" or "corresponding to reality" should be explained in terms of the accuracy with which the language user represents the world. To do so is to take the view that the concept of language "corresponding to reality" is itself an epistemic concept. I would be inclined to attempt such an explanation; if it were a plausible explanation, we could claim that if we cash in the metaphor of a person accurately/inaccurately representing the world, we are at the same time cashing in the metaphor of a person's language tightly/loosely fitting the world. That would be a nice economy: both of the metaphors we are inclined to use, when saying what is "unintelligible" in Fred's talk and thought, cashed in at one swoop. But it would lead us rather far afield; most of the issues one needs to examine in order to connect "correspondence with reality" with "representational accuracy" have little to do with the topic of confusion, and there are an awful lot of those issues. So I will leave it an open question whether we have come up with two unrelated metaphors to capture the sense in which the things Fred says are "unintelligible," or whether there is just one idea here, posing as two rather different ideas.
7. In his important essay "Meaning and Misconceptions," Anil Gupta considers a situation in which it is plausible to say that the users of a simple empirically incoherent language make "true" and "false" statements in the language "locally," even though a "global" rule of use conflicts with the perception-based criterion they rely upon. Gupta's proposal is that we think of the global rule as "selectively suspended." In the cases Gupta considers, the language-users meet with success in practice as a result of their local use of their globally incoherent language. This makes their use of language "truelike," and gives us a reason to find a consistent principle for ascribing truth. (Gupta does not use exactly this "local"/"global"/"selectively suspend"/"truelike" lingo.)

   If we focused on Fred's "Charley" talk during the intermittent and brief pe-

riods of time when, say, Ant A was scampering around before his eyes; if we cared mainly about what he said and did at those times, Gupta's analysis might fit Fred's case neatly. But as I have told the story, Fred evolves many of his "Charley" beliefs over time, as a result of encounters with both big ants, and does much of his talking "about Charley" in circumstances where he has no perceptual access to whichever big ant is up top in the ant colony at the moment he speaks. It is in circumstances like these that one has the strongest intuition that a person is confusing two things, and it is in circumstances like these that one has the least inclination to call the person's statements "true-like." It does seem right to attribute these intuitions to judgments by us that Fred is not using "Charley" language in a way that is conducive to success. I elaborate upon that idea in two different directions in Chapter 7 and in Part Five.

7. Calibration

1. This argument, and some of other arguments in this chapter, rest ultimately on the assumption that the (real) likelihood that a (real) likelihood will reflect the actual distribution in a class of occurrences gets higher as the class gets big. That is, I assume the conditions of the central limit theorem are fulfilled. I say this mainly because many epistemologists tend to complain against the applicability of such long-run considerations, on "Humean" grounds I do not fully grasp.
2. To see this intuitively, make the simplifying assumption that the ants do not influence each other at all with respect to having (to lacking) the relevant properties. Then imagine that Ant A has the symptoms on 1,000 occasions of sampling. Imagine further that half the time Ant B also has the symptoms. On 800 of the sampled occasions, Ant A has a cold. Of the 500 times both ants have the symptoms, Ant A has a cold on about 400 occasions. On the 500 occasions both ants have the symptoms, Ant B has a cold on about 400 occasions. But it is very likely that these are not exactly the same 400 occasions on which Ant A has a cold. So it is very likely that both Ant A and Ant B have colds on rather fewer than 400 of the 500 occasions when both have the symptoms. It takes both Ant A and Ant B having colds to make "Charley has a cold" supervaluationally true. So sentence (d), "Charley has a cold," probably is true on rather fewer than 400 of the 500 occasions when all of the sentences (a), (b), and (c) are true. That means that the conditional objective probability of (d) given the conjunction of (a) and (b) and (c) is rather less than 0.8.
3. Strictly, my assertions that Fred is poorly calibrated should be probabilistic. It is *likely* that he is. In order to avoid a bewildering iteration of probability modifiers, I make unmodified claims such as "Fred's judgment of the weight of the evidence (a), (b), and (c) for the conclusion (d) is wrong (high)." The argument is unaffected.

4. I am indebted to John MacFarlane for getting me to sort out the argument in the second half of this chapter.

8. Failure to Refer

1. See Robert Brandom, *Making It Explicit*.
2. If a philosopher believes the concept of Reference can be understood quite apart from understanding these other concepts, that philosopher owes an argument that she is using "refer" to express a language-world relation, rather than using it to represent intralinguistic anaphora.
3. When a distinction is made between "semantics" and "pragmatics," the concept "attempting to refer" falls on the side of pragmatics. That is all right; semantic concepts can be used in the articulation of nonsemantic concepts, just as the concept of entropy, drawn from physics, can be used to articulate the nonphysical concept "trying to raise the entropy of a rubber band" (easiest done by letting it snap if you have stretched it).
4. The paradigm of a simultaneous introduction of these interdependent concepts is, of course, Tarski's. See "The Concept of Truth in Formalized Languages," English translation by J. H. Woodger, in A. Tarski, *Logic, Semantics, Metamathematics* (Oxford: Oxford University Press, 1956). With enough relative-truth concepts, e.g., truth at a world or truth at a time, the remaining concepts of the cluster can be understood without explicitly marshaling the concept of Truth. See, for instance, Richard Montague, "English as a Formal Language," in *Formal Philosophy: Selected Papers of Richard Montague,* ed. with an introduction by R. H. Thomason (New Haven: Yale University Press, 1974).
5. This second step has come in for a good deal of abuse. "The man with the martini is my boss" says someone at a party, but her boss—the very man she is subtly gesturing toward—has a glass of water. See Keith S. Donnellan, "Reference and Definite Descriptions," *Philosophical Review* LXXV (1956), pp. 281–304. This seems to be a situation where the definite description is being used Referentially, but descriptive fit is weak. One might try to preserve the general idea of step two even in the face of such examples by substituting for "descriptive fit" some more general idea, such as that the speaker Refers to whatever object has a feature the speaker reasonably takes to be the feature expressed by the description. In the example cited, the speaker mistakes the attribute *having a glass of water* for the attribute *having a martini,* and it is reasonable of her to do so. Her Reference to her boss is secured by the fact that her boss has a glass of water (uniquely among things in her field of view). Of course we must not identify attributes with meanings if we take this line, since the speaker obviously does not mistake the meaning of the phrase "having a glass of water" for the meaning of the phrase "having a martini." My discomfort with the theory of descriptions does not lie in this direction, so let's stick with our canonical

formulation. (A note: When I speak of "using a definite description to Refer," I do not mean what Keith Donnellan means when he speaks of "referential [uses of] definite descriptions.")

6. A point, and a cute example, borrowed from John Wisdom, "Logical Constructions" (I), *Mind,* n.s. 40/158 (April 1941), pp. 188—112 (see p. 189).
7. Simple suggestion: a definite description "the F" can be used "distributively," so that the statement "The F is G" means "All Fs are G." It may be that there are such "distributive" uses of definite descriptions, and of other singular terms. But often, when a definite description is used to generalize, the generalization expressed by "The F is G" cannot be obtained by simply writing down "All Fs are G." For example, "The whale is rapidly disappearing from temperate waters" does not announce the departure of each individual whale. See also Richmond H. Thomason's discussion of the inference "The temperature is ninety, the temperature is rising, so ninety is rising," and other (possibly) kindred inferences in his "Home Is Where the Heart Is," in P. A. French, T. E Uehling, and H. K. Wettstein, ed., *Contemporary Perspectives,* pp. 209–219.
8. And as may be happening tacitly, in the Belloc example—a generalization containing the quantifier being assumed.
9. Peter Geach called attention to this phenomenon in "Intentional Identity," *Journal of Philosophy* 64 (1967). Walter Edelberg shows how very weird Geach sentences are from a semantic point of view in "A New Puzzle about Intentional Identity," *Journal of Philosophical Logic* 15 (1986), pp. 1–25. He gives a good general analysis, proposing a semantics, in W. Edelberg, "Intentional Identity," Ph.D. diss., University of Pittsburgh, 1984.
10. See Saul A. Kripke, *Naming and Necessity* (Cambridge, Mass.: Harvard University Press, 1980), for instance p. 56, or n. 33 on p. 79.
11. Here is why I append to Thesis K a reminder that there is a difference between knowing and knowing that one knows: someone might believe that whether or not a name-introducer has succeeded in her attempt to fix a Reference for the name **n,** she can know that if **n** exists, then **n** is the F, and she can know this merely by reflecting upon her attempt at Reference-fixing; but she cannot know that **n** is the F *(simpliciter)* without knowing that her act of Reference-fixing has succeeded. But that is wrong. If you affix the name "Jack" to the dog in the corner (so described), then you do know that Jack is the dog in the corner, you know it even if you are excessively paranoid and feel unsure whether you are only hallucinating a dog in the corner, leading you to doubt your knowledge. You do not know that you know, but you know.
12. It occurs to me that someone might be put off by my use in this argument of the turn of phrase "know that it is true," instead of plain "know"—on the grounds that it makes the argument too easy. If you have that reaction, replace the argument as given by this slightly longer alternate:

    If "the big ant" does not Refer to any big ant, then, according to the theory of descriptions, the sentence "Charley is the big ant" lacks a truth-value. And if "the big ant" Refers to an overabundance of big ants, then once again, accord-

ing to the theory of descriptions, the sentence "Charley is the big ant" lacks a truth-value. But if the sentence "Charley is the big ant" lacks a truth-value, then it is not the case that Fred knows that Charley is the big ant, since knowledge implies truth. This even holds "in the vicinity of" his attempt at Reference-fixing. By Thesis K it follows that Fred did not successfully introduce "Charley" into the language as a Referring term. Therefore, as Fred uses the name "Charley" it does not Refer at all.

9. How You Convince People

1. George Wilson, "Pronouns and Pronominal Descriptions—A New Semantical Category," *Philosophical Studies* 45 (1984), pp. 1–30. My argument in this chapter is heavily indebted to Wilson's important essay. See also his "Reference and Definite Descriptions," *Journal of Philosophy* 88 (1991), pp. 359–87. Jeffrey King discusses ways to elaborate this idea into a full-blown semantics in several papers. See his "Anaphors and Operators," *Philosophical Perspectives* 8 (Logic and Language), 1994; his "Pronouns, Descriptions, and the Semantics of Discourse," *Philosophical Studies* 51 (1987), pp. 341–63; and also his "Anaphora, Instantial Terms, and Pronouns," *Philosophical Studies* 61 (1991), pp. 239–65. This use of the term "parameter" is borrowed from Richmond H. Thomason, *Symbolic Logic; An Introduction* (New York: Macmillan, 1969). I am embellishing the term with a prefixed "E-" to distinguish parameters put into play by instantiating E-sentences from parameters put into play by instantiating A-sentences.
2. I continue to follow George Wilson's line of thought here, though the emphasis on parameters as "locally meaningful expressions" rather than "globally meaningful expressions" is a spin I borrow mainly from Mark Wilson. See his "Can We Trust Logical Form?," *Journal of Philosophy* 91 (1994), pp. 519–544. Robert Brandom has made very similar points using a somewhat different philosophical vocabulary. I do not know how long any of these writers would be willing to stay on my train, even though they built much of the engine.
3. This chapter evolved in the course of several discussions with Tamara Horowitz; she thought of some of it, I thought of some of it, and some of it we worked out together.

10. Trying to Predicate Existence

1. Immanuel Kant, *Critique of Pure Reason,* trans. Norman Kemp Smith (New York: St. Martin's Press, 1965), p. 505.
2. Richard Gale, "Existence, Tense and Presupposition," *Monist,* 50 (1966). I will alter Gale's argument just a little, so that it makes exactly the point I want it to make.
3. See n. 5 of Chapter 8.

4. Philosophers educated to interpret E-sentences as existentials sometimes have trouble hearing the first conjunct of (10), "There is no old man Susan daydreams about," as meaning just "Susan does not daydream about an old man," but this is exactly what it does mean. If you are one of those philosophers whose ear has been corrupted, it will help to think about a string of sentences such as: (i) The old man Susan daydreams about must be on her mind again; (ii) Susan spends too much time daydreaming about the old man she daydreams about; (iii) especially since the old man Susan daydreams about is imaginary; (iv) that is to say, the old man Susan daydreams about does not exist. And then throw yourself a curve by adding: (v) of course there is no old man Susan daydreams about.
5. Observe that if this is your theory of choice, you will have interpreted some existentials as genuine ascriptions of the property "being based on something real" (or denials of such). That is a pretty good candidate for what it would be to ascribe (or deny) existence to a subject, leading to a theoretical objection to (I). The subjects of this predication are peculiar entities. But as the theory would have it, they can possess such properties as "being an old man Susan daydreams about." This seems to make them more nearly particulars than predicables, leading to a theoretical objection to (II). Those are quick estimates, but they will suffice to locate the position in the traditional debate.

   And how does the question whether existence is a real predicate look by the lights of our ungrammaticality explanation of the "contradiction" in (10)? Here is a picture that may be helpful: think of the language we use to talk about Susan as a two-sorted Fitchish quantifier language. Write the logically particular sentence "There is an old man Susan daydreams about" as "[Σx][SDx]," where "[Σx]" is a "particular" quantifier assumed to have the ordinary Fitch-system syntax. For concreteness think of it as objectually interpreted, but in an inflated domain. Let "(Ex)" be an existential quantifier, an objectually interpreted quantifier with an "ordinary" domain of real things. (Arrange things so that "(Ex)(SDx)" is not a valid consequence of "[Σx][SDx].") Finally, write "The old man Susan daydreams about does not exist" as "-(Ex)(x=α)," where α is an E-parameter put into play by instantiating "[Σx][SDx]." The E-parameter α is to be the analogue of the definite description "the old man Susan daydreams about"; this amounts to treating the logically particular generalization "There is an old man Susan daydreams about" as the only available antecedent for the anaphor "the old man Susan daydreams about," which is unrealistic, but for simplicity assume it.

   If the E-parameter α were a singular term, then the sentence "-(Ex)(x=α)" would be a singular predication with a quantificationally complex predicate. We might conclude that "exists" is a real predicate here—a predicate even in the deep structure. But parameters like α are not singular terms at all, so "-(Ex)(x=α)" is not a singular subject-predicate sentence (cf. George Wilson, "Pronouns").

6. When I mentioned "respectable philosophical goals" that might be reached by adopting Option 4, I had in mind the thought of unifying the semantics of confused language with the semantics of what can be called "attributions of *joint,* or *collective,* thought"—e.g., when one speaks as though two or more people have views about the same unreal object. I once made a stab at setting out some of the issues, in Joseph L. Camp, Jr., "Why Attributions of Aboutness Report Soft Facts," *Philosophical Topics* 16 (1988), pp. 5–30.

11. Explicating

1. More precisely, that was the upshot of the calibration argument of Chapter 7; the point of Chapters 8, 9, and 10 was to undermine a different and prima facie attractive argument for the same conclusion. I do not know of a good argument to the effect that Fred's "Charley" talk and "Charley" thought cannot be Truth-valued, other than the argument of Chapter 7. Indeed, the calibration considerations that drive the argument of Chapter 7 suggest that other arguments are not to be had. Here is one illustration:

   Someone might wonder whether an argument can be devised which shows that the Reference/Truth/Truth-of cluster of semantic concepts is applicable to chunks of Fred's thought that contain "Charley" sentences in addition to other, less bizarre, sentences, except that the "Charley" sentences themselves happen to be neither True nor False. For example, why not adopt a "causal" conception of Reference, according to which a name as used by a given speaker Refers to that item which is the dominant cause of those of the speaker's beliefs (transparently) containing the name? The idea needs to be further elaborated, but does it not have promise as the semantic component of an explanation of Fred's confusion, and without junking the classical semantic concepts? It just gives those concepts a causal spin. One might argue that Fred's "Charley" beliefs do not have a dominant cause, since they arise more or less equally from encounters with Ant A and encounters with Ant B. So Fred's "Charley" beliefs are, as it happens, neither True nor False. But other beliefs of Fred's that may well "mix" with his "Charley" beliefs in inferences will have Truth-values. For example, Fred might reason that Charley can lift Sally the Stick, and Sally the Stick is heavy, so Charley can lift a heavy thing. Presumably Fred's "Sally the Stick" beliefs do have a dominant cause, so we will maintain that Fred uses the name "Sally the Stick" to Refer (to a certain stick). Thus not all of the terms occurring in a chunk of "confused thought" need lack Referents. We will be able to retain the intuition that Fred can Refer to Sally even if in the same breath he makes a confused "Charley" claim. We will have placed confused beliefs and ordinary Truth-valuable beliefs on the same scale, as it were. Is that not a desirable outcome?

   It is, but the "causal" diagnosis of confusion is wrong. Set things up like this: Fred has encountered Ant A fourteen hundred times, but he has only encoun-

tered Ant B four hundred times. This has happened by "accident." That is, we cannot predict, statistically, that these ratios will persist in Fred's "life with Charley"; the likelihoods are that Fred's future "Charley- encounters" will cluster around 0.5 Ant A/0.5 Ant B, so Fred is at high epistemic risk in his "Charley"-reasoning. Intuitively, does Fred use the name "Charley" to Refer to Ant A, the object by which his "Charley" beliefs are dominantly caused? No. Fred is still confused. Now make it 2,000/1 in favor of Ant A. But stipulate, again, that this ratio is an accident, a fluke, so that the 2,000/1 ratio is not projectible. Once again, Fred is confused. The reason a dominant-cause analysis of the Reference of "Charley" seems attractive on first sight is that we correctly infer from the information that Fred's "Charley"-beliefs are about equally caused by encounters with Ant A and encounters with Ant B, so that a continued equal likelihood is projectible. This inference is correct when the only information we have that is statistically relevant to the question whether Fred is at epistemic risk if he (now) goes about framing "Charley"-beliefs is our information about ratios of encounters in the past. But when we have in addition some information to the effect that the ratios of encounters in the past are statistically unrepresentative (the idea I summarized above by calling the past ratio "an accident"), then the actual past ratio of Ant A/Ant B encounters does not influence our judgment of confusion; what does influence it is the riskiness of Fred's inferences. So we can pry off the element of riskiness from the element of causal origins, and we can see that what matters in our judgment of confusion is not per se that Fred's "Charley"-beliefs have a mix of Ant A and Ant B causal origins; causal origins matter only insofar as they are evidence for riskiness assignments.

In fact I believe this line of thought is a refutation of causal theories of Reference quite generally, but that would take us far beyond the task at hand. The point here is that the argument of Chapter 7 picks up the idea of epistemic riskiness, abstracted from all other associated features, such as mixed causal origins. I cannot prove that the element of epistemic riskiness, spelled out in terms of calibration, is bound to keep popping up like Austin's frog at the bottom of the beer mug, but I am convinced it will. (I am indebted to a helpful email discussion with John MacFarlane's Fall 2000 graduate seminar at U.C. Berkeley.)

2. Marvin and his judgments of thermal relations appeared in Chapter 2, only to drop out of sight thereafter. He will be reincarnated as an entire civilization before long, so it is understandable that he would decide to rest.
3. Carnap did not describe explication this way, but I believe he should have and would have if he had worked with the idea of confusion rather than the idea of vagueness. That is not quickly argued, though, so if you disagree, just take "explicate" as having a stipulated and inauthentic meaning here.
4. A parenthetical observation: the state Fred would be in after we had taught him all of these things would be correctly described by saying he uses the name "Charley" equivocally. But he would be able to "access" his equivocation the

way a person who uses "bank" equivocally can access her equivocation. That is, he would have a supplementary linguistic practice—using the names "Ant A" and "Ant B" instead of "Charley"—and he would understand how to translate back and forth. If he committed fallacies of equivocation under these conditions, there would be no basis for inferential charity toward him: these are not the sort of inferences a completely logical thinker makes. *Therefore he would no longer be confused*. So we should not mistake the circumstances just described for the circumstances that exist when we do attribute confusion to Fred. When we do attribute confusion to him, he has no "fallback" linguistic practice. In that case, if we treat the name "Charley" as equivocal, we will force ourselves to classify as fallacies of equivocation many arguments he gives which we think he is being completely logical in giving. (See Chapter 5.)

12. Good Advice

1. As I have just told the story, "we" do not believe in one of the "things" at issue—specific heat—but there is no reason why "we" ought not to believe in temperature. This suffices to make my point, but if the reader prefers, she can complicate the story so that "we" take the Successors to have mistakenly believed in temperature also. It will be a messier story, since quantum and classical "conceptions" of temperature do not (really) diverge in the neat way the two models of specific heat (really) do.
2. Less tendentiously: I do not wish to defend a philosophy of mathematics that assimilates "mathematical facts" to facts of nature.
3. Jamie Tappenden nicely brings out the magnitude of this shift in J. Tappenden, "Extending Knowledge and 'Fruitful Concepts': Fregean Themes in the Foundations of Mathematics," *Noûs* 29 (1995), pp. 427–67, though he is discussing Frege's take on it.
4. Nuel Belnap, "A Useful Four-valued Logic," in Anderson, Belnap, and Dunn, *Entailment,* pp. 506–541. The interpretation I am putting on the semantic values is close to Belnap's "F.B.I. files" interpretation, although he was not especially concerned to fit that interpretation to confused language.
5. The authorities one appeals to in defining the properties **Y, N, Y&N,** and **?** for a given example of confusion might be the confused person herself, later, wearing two different hats—a "Sam hat" and a "Sal hat." Consider looking back on the parking lot experience described in Chapter 1. You can play Sam and Sal at once, alternately using "this car" to denote the car directly before you-at-that-earlier-time, or the car across the parking lot from you-at-that-earlier-time.

13. How Fred Should Think

1. Or rather, this is the ideal: we want a logically valid argument to have these features. If we knew how to do it, we might broaden the scope of our explication

of "follows-from" so as to include strong nonmonotonic arguments that are less than valid but very reasonable of Fred to make.

2. A theme reiterated from Chapter 7.
3. The most attractive way of extending the semantics to quantified sentences is a substitution interpretation. But think about this: intuitively, in a substitution semantics the quantified sentence (x)(Fx) is supposed to work "like an infinite conjunction" of the instances Fa, Fb, and so on, for all suitable substitution instances a, b, and so on, supposing there to be infinitely many substitutable terms. For many predicates F it is unreasonable to think Sam and Sal will have found a way to investigate each instance Fa, Fb, etc., and it is unreasonable to think there is an answer, counterfactually, what they would say in all of those cases if they did investigate. In such a case, some instances of the generalization should get the value **?**. A glance at the value table for conjunction shows that in such cases the generalization (x)(Fx) should itself get the value **?**—continuing to press the analogy between a substitutional universal generalization and a big conjunction.

   This raises a question how we can understand Fred's instantial inductive reasonings. The superficially plausible approach would be to say that when Fred himself confirms a bunch of instances Fa, Fb, etc., what he does is "gather evidence that each is **Y**" though he would not describe it that way. But if the generalization (x)(Fx) is **?** in many such cases, there will be many failures, perhaps too many, of the inductive principle "From a number of **Y** instances infer that you have a **Y** universal generalization." The instances that shoot down Fred's inductions may well involve terms Sam and Sal use to denote run-of-the-mill ants or pebbles or atoms, not the term "Charley." This is a serious worry, but I'll leave it at that.
4. Belnap, in Anderson, Belnap, and Dunn, *Entailment,* v. II, p. 517.
5. (1) does not imply (2). Imagine a "semantics" that defined "validity" as P-preservation, for some oddball property P. It could happen that (a) every argument we think of as "logical" preserves P; and (b) every P-preserving argument is one we would think of as "logical"; but (c) being P-preserving does not have anything to do with the concept "follows-from" except that the two concepts necessarily coincide in extension. (Model theorists, being mean-spirited by nature, have devised various messy examples, none of which I will repeat here.)
6. Anderson, Belnap, and Dunn, *Entailment,* pp. 519–520.
7. I often write as though there were a single, hard-edged, set of "pre-theoretic intuitions" concerning what follows from what. This is not true of me, and I believe was not true of me a long time ago, when my "intuitions" were much less theory-laden than they are today. I imagine it is not true of anyone. "Follows-from" is vague; one can find candidate inferences of which one simply cannot tell whether they are, intuitively, valid. This is so even after one has settled on the background epistemological considerations one deems important. Professional philosophers may tend to agree rather more than nonphiloso-

phers in their validity intuitions, but most professional philosophers have had instruction in logic, and that instruction nearly always has been in classical logic. In that condition, its hard to tell a "pretheoretical" intuition from a homework answer.

8. A helpful discussion of disjunctive syllogism in the kind of logic we want to apply to Fred is Anderson, Belnap, and Dunn, *Entailment,* pp. 488–505.
9. Anil Gupta, John MacFarlane, and Mark Wilson helped me correct several mistakes in an earlier draft of this chapter, some of which were spill-over mistakes from the chapter before. Much indebted.

14. Semantic Self-Awareness

1. I follow G. H. Hardy's account in his *Divergent Series,* 2nd ed. (Providence: Chelsea Publishing Company, 1991). Hardy says that Leibniz had worked out his views about this series as early as 1713 (ibid., p. 8). I have not confirmed that.
2. It is mainly that we do not know whether the real Leibniz ever had this sequence of thoughts. It is as though he had them, I believe, but I do not want to spend time trying to defend that.
3. Mark Wilson has argued convincingly that situations like the one Leibniz* finds himself in are not uncommon in the history of science. Wilson's emphasis is on the need for research to progress in spite of the partial unintelligibility of the theory-language in question, in the hope that later researchers will provide what amounts to a semantic ratification. He has in mind situations in which researchers are meeting with "success" of various kinds by thinking in their semantically dubious terms, or at least where they have an intuitive feel for how to steer clear of the risks. Leibniz* is not in that fortunate position: if we consider the relevant subject-matter to be alternating infinite sums (rather than infinite sums in general), Leibniz* does not even know how to tell if he is getting "fruitful" results. See Mark Wilson, "Can We Trust Logical Forms?".

   I could not have thought any of this through had I not studied Wilson's work first, and had several instructive conversations with him.
4. Evidently the real Leibniz did not tinker quite as much as Leibniz*, since the only alternating series to which Leibniz applied this neat—if a tad limited—conception of convergence was the series s itself. It remained for others to generalize the idea; J. Bernoulli, for one, and Cesaro. The story is well told by Hardy, in *Divergent Series.*
5. See ibid. for the main mathematical part of the arguments. I am just restating the conclusions to suit my philosophical themes.

15. Two Charleys

1. See Anil Gupta, *The Logic of Common Nouns: An Investigation in Quantified Modal Logic* (New Haven: Yale University Press, 1980).

2. The psychological explanation we gave would not be of "the fact that Fred cannot distinguish Ant A from Ant B in thought"—one does not give psychological explanations of semantic "facts"; of the content of semantic positions. We would be explaining why Fred did not register the expected bafflement, and why he did not feel driven to adjust his way of talking.

16. Young Newton

1. If the innovator is on the right side of history, or the "distinction" she claims to have made logically visible is especially fruitful, people will keep on employing the forms of argument she has pioneered and pray for a semantic paladin to come along and semantically ratify the damn things. When the story plays out along these lines, we have a special case of the type of historical phenomenon Mark Wilson describes in his "Can We Trust Logical Forms?"
2. Probably Berkeley usually has in mind Newton's formulation of the calculus in his later treatise on the quadrature of curves. Some people believe Newton's stark reliance on "infinitesimal" reasoning in his early work was supplanted in his more mature work by "limit" reasoning. But it is anachronistic to find even nineteenth-century limit reasoning in the total absence of any attendant topological concepts (e.g., neighborhood) and in the absence of even a modicum of respect for quantifier order. Not that I have much right griping about anachronism.
3. George Berkeley, *De Motu and The Analyst,* ed. and trans. by Douglas Jesseph (Norwell, Mass.: Kluwer Academic Publishers, 1992), p. 200.
4. Ibid., p. 201.
5. And indeed he speaks of the difficulty one would have framing any "Idea or Notion" to do this work. In his mature philosophy he used the term 'Notion' to mean roughly "concept of something of which you cannot form an image." So picturability is not the issue.
6. Berkeley, *De Motu,* p. 195.
7. Ibid., p. 205. See also all of *Analyst* sections 31, and 42–44 (ibid., p. 194 and pp. 204–205).
8. An excellent introduction is John L. Bell, *A Primer of Infinitesimal Analysis* (New York: Cambridge University Press, 1998).
9. This is not, in particular, because "infinitesimals" are being modeled. Double negation fails in the "topos logic" for categories other than Set (and, of course, in intuitionism).
10. The other widely studied construction of infinitesimals that is relevant to semantically interpreting Newton is Abraham Robinson's. I have a belief that the concepts of sameness and difference are altered by the compactness arguments underlying Robinson's construction as much as they are in the category-theorists' construction. But I still cannot think of a satisfactory argument for that belief. Other modern constructions of "the infinitely small," for instance the delta-distribution construction of infinitely brief "pulses," probably

have application in interpreting certain parts of Newton's dynamical thinking, but I have not figured out whether a sameness/difference point can be made in connection with that construction (a clear introduction to delta-distribution ideas in dynamics is Ram P. Kanwal, *Generalized Functions; Theory and Technique,* 2nd ed. (Basel: Birkhauser, 1998).

18. The Theory of Ideas

1. In the *Third Meditation;* see J. Cottingham, R. Stoothoff, and D. Murdoch, trans. and ed., *The Philosophical Writings of Descartes,* vol. II, p. 25.
2. See note 6 of Chapter 2.
3. "Materially" does not mean "physically." It means something like "made of the stuff that is to be found there." In churches the stuff is stone and wood. In minds the stuff is nonphysical, but stuff all the same.
4. This is a weaker claim than I really want to make. What I really want to claim is that I am ready to accept that Descartes believed he understood what it would be for (1) to be true, because I myself "almost understand" what it would be for (1) to be true. Since I cannot explain what I mean by "almost understand," I have copped out and put the weaker argument in the text, reserving my real argument for a footnote where it is less embarrassing.
5. David Hume, *A Treatise of Human Nature,* ed. L. A. Selby-Bigge (Oxford: Oxford University Press, 1888), p. 94.
6. If you have one "Charley" idea for two ants, or if you have one idea of "cold" for two different objects—a certain micromechanical state of matter and the type of sensation you have when, say, you grab a snowball—then the "intentionality is identity" slogan has been pushed past the breaking point. Descartes had no metaphysical alternative, and had to scratch out his meaning as best he could (as we saw in Chapter 2).
7. A good survey is Robert Pasnau, *Theories of Cognition in the Later Middle Ages* (New York: Cambridge University Press, 1997).
8. Berkeley's reactionary suggestion was that the "particulars" that can exist in the mind really are just batches of properties, so the traditional theory got the mechanism of attachment right after all. He also believed the properties in the property-bundles are experiential contents, but this was a bit of empiricist dogma easily separated from the main point.
9. I believe a mark of seventeenth-century science and philosophy was a diminished respect for distinctions of reason, and I believe this was liberating. One trembles at the reassertion of the importance of "grammar" in twentieth-century analytic philosophy. I won't take time to defend any of those opinions here.
10. The opinion was at least partly a residuum of ancient atomism. For a helpful study see Andrew Pyle, *Atomism and Its Critics: From Democritus to Newton* (Bristol, U.K.: Thoemmes Press, 1997). I do not suggest that the "atomism" of the mechanical philosophy was just a rehash; plainly it was not. The

"atomic" items of early modern mechanism were minimal "natural" machine components, a different concept entirely from the bare concept of a smallest part.

11. Or—likeliest, I suspect—a terrain map of the "new" type that did not attempt "lifelike portrayal" of terrain features or structures but did have a scale permitting one to read spatial relations and dimensions from the map to the ground, and also obeyed the convention that numbers of map-markers reflected numbers of wordly items (trees in the orchard, houses in the city, etc.). The cardinal feature of maps drawn in this "new" style was that the map contained accurate primary-quality information encoded in pictorial form. The cost was that such maps no longer looked very much like the scene depicted. Maps drawn in the older fashion showed one a sort of bird's-eye view out over the landscape; they were rich in secondary-quality information though poor in primary-quality information (you can see that the thing is a church, even that it is has a graceful white steeple, but it would be a mistake to infer that the number of houses portrayed as in the town is the number of houses actually in the town). By the time John Locke started writing in the philosophy of mind and the philosophy of language, many excellent examples of both types of map were floating around. Locke must have seen a good sample, since people with a little money tended to own and display rather more maps than they do now. See P. D. A. Harvey, *Maps in Tudor England* (Chicago: University of Chicago Press, 1993).
12. "A circle and a square are the same," Locke said, ". . . in the mind and in the manna." (Locke, *Essay,* p. 138). In the same breath he claimed that the whiteness one finds in one's idea of manna is on the same footing as the nausea one experiences upon consuming the stuff. (I gather that manna was stuff you swallowed to make yourself sick. Evidently it was sweet and white. Locke, the sometime physician, probably had crates of it around.) If the microstructure of the manna is responsible for its effects on a human, and the only way these effects can be produced is by mechanical interactions between manna-microstructure and stomach-microstructure, then an idea of manna, even a "blow-up" of a tiny region of manna, would be no good for understanding these interactions if it did not accurately represent mechanically efficacious features, like shape.

19. Making Category Mistakes

1. Back in Chapter 2.
2. Locke, *Essay* IV, VII, 4, p. 592.
3. Locke, *Essay* II, XXIX, 5, p. 364.
4. Locke, *Essay* IV, VII, 10, p. 597.
5. Ibid., 4, pp. 592–593. Knowledge of these identities and differences is trivial, Locke claims—taking a shot at Aristotelians and Cartesians who elevate some instances of such knowledge to the rank of metaphysical first principles. That is

not to say it is useless knowledge. By stringing together enough bits of it, Locke believed, a clever person can discover identities and differences between the extreme terms of the chain. Those identities and differences often are unobvious and important.

6. Recall Chapter 2.
7. This is the long version of the anti-Cartesian empiricist argument I sketched in Chapter 2.
8. Notice that "nonextensional" cannot have its usual meaning once we adopt an epistemic semantics. There is no longer a distinction to be made between connectives which express truth-functions and connectives which do not. There is, however, a distinction to be made between connectives which express value-functions and connectives which do not. Much, though not all, of what is logically annoying about non-truth-functionality is a result of its being a species of non-value-functionality. I will be very surprised if the connective "Herb knows that . . ." or the connective "Sally clearly perceives that . . ." or the connective "It is possible that . . ." turn out to be value-functional when the details of an extended Belnapian semantics are worked out.
9. Recall the discussion in Chapter 1. The "didn't anticipate later developments" rationale for deploying the value "neither-true-nor-false" is not the same as either of the two rationales mentioned in Chapter 1.
10. Example: with a scanning tunneling microscope—an essentially quantum mechanical device—one can make a photograph of the surface of a piece of glass (silica) that shows the silicon atoms as a neat array of fuzzy balls with the occasional big fuzzy ball of an oxygen atom imbedded. Having seen a few such photographs, I can picture the scene easily. Do I have an "idea of the atomic plane of silica," by the lights of Locke's version of the theory of ideas? I can form rather a good visual image (the photographs are memorable). But my basis for *taking the image to be of* the atomic plane of silica is that persons I believe to be qualified surface chemists have told me so. This trust-based rule of acceptance is not on the traditional empiricist menu. The surface chemists themselves have different reasons for taking their own (similar) visual images to be of the atomic plane of silica, but those reasons, being deeply quantum-mechanical, barely resemble anything Locke ever imagined could be a "reason." Then again, Locke evidently thought we would have different ideas if we had microscopical eyes—though presumably he had in mind late seventeenth-century microscopes. I bet you won't settle on an answer to the question whether I have the idea. "Neither-true-nor-false" seems about right.
11. If you have managed to avoid the exercise until now, form a visual image *of green* (*simpliciter*). What did you do? Picture a splash of green more or less covering all of your visual field? Fine. Now, does an image of an array of green discs have that splash-of-green image as a part? No. And so forth.
12. Since we have these examples of idea-talk in front of us, this is a good time to say something about a suggestion Wilfrid Sellars makes in "Empiricism and the Philosophy of Mind," ch. 5 of his *Science, Perception, and Reality* (Lon-

don: Routledge and Kegan Paul, 1963), pp. 127–196, esp. pp. 190–196. The suggestion is that we look at the theory of ideas in its empiricist incarnation this way: you start with sensory awarenesses, and then augment the domain of properties an awareness can possess, and the legitimate inferences one can draw concerning awarenesses, by analogy with the properties of, and the inferences one can draw concerning, a bunch of colored shaped wafers. This new, analogically introduced pattern of inferences lets you get the effect of ideas "being objects" while they still "really are" awareness-states.

This is hardly a charitable reconstruction. Suppose that before making the analogical extension, one can give the following sound argument (we are remaining semantically classical, so we can speak of "soundness"):

(A) Susan's visual representation of an array of five green discs is an awareness

(B) No awareness resembles any array of wafers geometrically or topologically.

So,

(C) Susan's visual representation of an array of five green discs does not resemble the array of wafers W with respect to geometric/topological property P.

We had better not extend the "logic of awarenesses" analogically so as to admit a parallel inference to

(D) Susan's visual representation of an array of five green discs *does* resemble the array of wafers W with respect to geometric/topological property P

or we have not got a consistent extension. Presumably one insures consistency by flagging the predicate of (D) thus:

(D_o) Susan's visual representation of an array of five green discs *does* resemble[object] the array of wafers W with respect to geometric/topological property P,

while flagging the predicate of (C) thus:

(C_a) Susan's visual representation of an array of five green discs does *not* resemble[act] the array of wafers W with respect to geometric/topological property P.

Or in some other fashion distinguish analogical uses of predicates from "primitive" uses. This has the same effect as distinguishing two senses of the term "idea." Arguments we do not wish to interpret as trivially invalid are trivially invalid; fallacies of equivocation.

Index