

SAUL A. KRIPKE

NAMING AND NECESSITY¹

Lectures Given to the Princeton University Philosophy Colloquium

LECTURE I: JANUARY 20, 1970

I hope that some people see some connection between the two topics in the title. If not, anyway, such connections will be developed in the course of these talks. Furthermore, because of the use of tools involving reference and necessity in analytic philosophy today, our views on these topics really have wide-ranging implications for other problems in philosophy that traditionally might be thought far-removed, like arguments over the mind-body problem or the so-called 'identity thesis'. Materialism, in this form, often now gets involved in very intricate ways in questions about what is necessary or contingent in identity of properties – questions like that. So, it is really very important to philosophers who may want to work in many domains to get clear about these concepts. Maybe I will say something about the mind-body problem in the course of these talks. I want to talk also at some point (I don't know if I can get it in) about substances and natural kinds.

The way I approach these matters will be, in some ways, quite different from what people are thinking today (though it also has some points of contact with what some people have been thinking and writing today, and if I leave people out in informal talks like this, I hope that I will be forgiven).² Some of the views that I have are views which may at first glance strike some as obviously wrong. My favorite example is this (which I probably won't defend in the lectures: for one thing it doesn't ever convince anyone). It is a common claim in contemporary philosophy that there are certain predicates which, though they are in fact empty – have null extension – have it as a matter of contingent fact and not as a matter of any sort of necessity. Well, *that* I don't dispute; but an example which is usually given is the example of *unicorn*. So it is said that though we have all found out that there are no unicorns, of course there *might* have been unicorns. Under certain circumstances there *would* have been

Exam
of us
a la
view.
Unic
don't
could
be t
1.0
5.0

Note: For addenda to this paper see pp. 763-769 in this volume.

*Davidson and Harman (eds.), Semantics of Natural Language, 253-355. All rights reserved
Copyright © 1972 by Saul A. Kripke*

unicorns. And this is an example of something I think is not the case. Perhaps according to me the truth should not be put in terms of saying that it is necessary that there be no unicorns, but just that we can't say under what circumstances there would have been unicorns. Further, I think that even if archeologists or geologists were to discover tomorrow some fossils conclusively showing the existence of animals in the past satisfying everything we know about the unicorns from the myth of the unicorn; that would not show that there were unicorns. Now I don't know if I'm going to have a chance to defend this particular view, but it's an example of a surprising one. (I actually gave a seminar in this institution where I talked about this view for a couple of sessions.) So, some of my opinions are somewhat surprising; but let us start out with some area that is perhaps not as surprising and introduce the methodology and problems of these talks.

The first topic in the pair of topics is naming. By a name here I will mean a proper name, i.e. the name of a person, a city, a country, etc. It is well known that modern logicians also are very interested in definite descriptions: phrases of the form 'the x such that ϕx ', such as 'the man who corrupted Hadleyburg'. Now, if one and only one man ever corrupted Hadleyburg, then that man is the referent, in the logician's sense, of that description. We will use the term 'name' so that it does *not* include definite descriptions of that sort, but only those things which in ordinary language would be called proper names. If we want a common term to cover names and descriptions we may use the term 'designator'.

It is a point, made by Donnellan,³ that under certain circumstances a particular speaker may use a definite description to refer, not to the proper referent, in the sense that I've just defined it, of that description, but to something else which he wants to single out and which he thinks is the proper referent of the description, but which in fact isn't. So, you may say "The man over there with the champagne in his glass is happy", though he actually only has water in his glass. Now, even though there is no champagne in his glass, and there may be another man in the room who does have champagne in his glass, the speaker *intended* to refer, or maybe, in some sense of 'refer', *did* refer, to the man he thought had the champagne in his glass. Nevertheless I'm just going to use the term 'referent of the description' to mean the object uniquely satisfying the conditions in the definite description. This is the sense in which it's been

Names have denotation but
no connotation.

used in the logical tradition. So, if you have a description of the form 'the x such that ϕx ', and there is exactly one x such that ϕx , that is the referent of the description.

Now, what is the relation between names and descriptions? There is a well known doctrine of John Stuart Mill, in his book *A System of Logic*, that names have denotation but not connotation. To use one of his examples, when we use the name 'Dartmouth' to describe a certain locality in England, it may be so called because it lies at the mouth of the Dart. But even, he says, had the Dart (that's a river) changed its course so that Dartmouth no longer lay at the mouth of the Dart, we could still with propriety call this place 'Dartmouth', even though the name may suggest that it lies at the mouth of the Dart. Changing Mill's terminology, perhaps we should say that a name such as 'Dartmouth' *does* have a 'connotation' to some people, namely, it *does* connote (not to me - I never thought of this) that any place called 'Dartmouth' lies at the mouth of the Dart. But then in some way it doesn't have a 'sense'. At least, it is not part of the *meaning* of the name 'Dartmouth' that the town so named lies at the mouth of the Dart. Someone who said that Dartmouth did not lie at the Dart's mouth would not contradict himself.

It should not be thought that every phrase of the form 'the x such that $F-x$ ' is always used in English as a description rather than a name. I guess everyone has heard about the Holy Roman Empire, which was neither holy, Roman nor an empire. Today we have the United Nations. Here it would seem that since these things can be so-called even though they are not Holy Roman United Nations, these phrases should be regarded not as definite descriptions, but as names. In the case of some terms, people might have doubts as to whether they're names or descriptions; like 'God' - does it describe God as the unique divine being or is it a name of God? But such cases needn't necessarily bother us.

Now here I am making a distinction which is certainly made in language. But the classical tradition of modern logic has gone very strongly against Mill's view. Frege and Russell both thought, and seemed to arrive at these conclusions independently of each other, that Mill was wrong in a very strong sense: really a proper name, properly used, simply was a definite description abbreviated or disguised. Frege specifically said that such a description gave the sense of the name.⁴

Is 'God' a name or a description?

Answer: Clearly a name.

Mill
Names
have
denotation
no connotation

Frege
Name
vs description
cannot

Now the reasons against Mill's view and in favor of the alternative view adopted by Frege and Russell are really very powerful; and it is hard to see – though one may be suspicious of this view because names don't seem to be disguised descriptions – how the Frege-Russell view, or some suitable variant, can fail to be the case.

Let me give an example of some of the arguments which seem conclusive in favor of the view of Frege and Russell. The basic problem for any view such as Mill's is how we can determine what the referent of a name, as used by a given speaker, is. According to the description view, the answer is clear. If 'Joe Doakes' is just short for 'the man who corrupted Hadleyburg', then whoever corrupted Hadleyburg uniquely is the referent of the name 'Joe Doakes'. However, if there is *not* such a descriptive content to the name, then how do people ever use names to refer to things at all? Well, they may be in a position to point to some things and thus determine the references of certain names ostensively. This was Russell's doctrine of acquaintance, which he thought the so-called genuine or proper names satisfied. But of course ordinary names refer to all sorts of people, like Walter Scott, to whom we can't possibly point. And our reference here seems to be determined by our knowledge of them. Whatever we know about them determines the referent of the name as the unique thing satisfying those properties. For example, if I use the name 'Napoleon', and someone asks, "To whom are you referring?", I will answer something like, "Napoleon was emperor of the French in the early part of the nineteenth century; he was eventually defeated at Waterloo", thus giving a uniquely identifying description to determine the referent of the name. Frege and Russell, then, appear to give the natural account of how reference is determined here; Mill appears to give none.

There are subsidiary arguments which, though they are based on more specialized problems, are also motivations for accepting the view. One is that sometimes we may discover that two names have the same referent, and express this by an identity statement. So, for example (I guess this is a hackneyed example), you see a star in the evening and it's called 'Hesperus'. (That's what we call it in the evening, is that right? – I hope it's not the other way around.) We see a star in the morning and call it 'Phosphorus'. Well, then, in fact we find that it's not a star, but is the planet Venus and that Hesperus and Phosphorus are in fact the same.

Meaning that names are descriptive
helps us understand the sense of
Hesperus = Phosphorus
NAMING AND NECESSITY
Cicero - Tully 257

So we express this by 'Hesperus is Phosphorus'. Here we're certainly not just saying of an object that it's identical with itself. This is something that we discovered. A very natural thing is to say that the real content [is that] the star which we saw in the evening is the star which we saw in the morning (or, more accurately, that the thing which we saw in the evening is the thing which we saw in the morning). This, then, gives the real meaning of the identity statement in question; and the analysis in terms of descriptions does this.

Also we may raise the question whether a name has any reference at all when we ask, e.g., whether Aristotle ever existed. It seems natural here to think that what is questioned is not whether this *thing* (man) existed. Once we've got the thing we know that it existed. What really is queried is whether anything answers to the properties we associate with the name - in the case of Aristotle, whether any one Greek philosopher produced certain works, or at least a suitable number of them.

Et
from
to us
that
a de
ion.

It would be nice to answer all of these arguments. I am not entirely able to see my way clear through every problem of this sort that can be raised. Furthermore, I'm pretty sure that I won't have time to discuss all these questions in these lectures. Nevertheless, I think it's pretty certain that the view of Frege and Russell is false.⁵

F-R
false

Many people have said that the theory of Frege and Russell is false, but, in my opinion, they have abandoned its letter while retaining its spirit, namely, they have used the notion of a cluster concept. Well, what is this? The biggest problem for Frege and Russell, the one which comes immediately to mind, is already mentioned by Frege himself. He said,

Probl
Man
dis
for
name

In the case of genuinely proper names like 'Aristotle' opinions as regards their sense may diverge. As such may, e.g., be suggested: Plato's disciple and the teacher of Alexander the Great. Whoever accepts this sense will interpret the meaning of the statement 'Aristotle was born in Stagira', differently from one who interpreted the sense of 'Aristotle' as the Stagirite teacher of Alexander the Great. As long as the nominatum remains the same, these fluctuations in sense are tolerable. But they should be avoided in the system of a demonstrative science and should not appear in a perfect language.⁶

So, according to Frege, there is some sort of looseness or weakness in our language. Some people may give one sense to the name 'Aristotle', others may give another. But of course it is not only that; even a single speaker when asked "What description are you willing to substitute for the name?" may be quite at a loss. In fact, he may know many things about him; but any particular thing that he knows he may feel clearly

'the man such that such and such and such and such', or even maybe with the family (if something can be synonymous with a family), the family or the single description is what is used to determine to whom some-one is referring when he says 'Walter Scott'. Of course, if when we hear his beliefs about Walter Scott we find that they are actually much more nearly true of Salvador Dali, then according to this theory the reference of this name is going to be Mr. Dali, not Scott. There are writers, I think, who explicitly deny that names have meaning at all even more strongly than I would but still use this picture of how the referent of the name gets determined. A good case in point is Paul Ziff in his *Semantic Analysis*, who says, very emphatically, that names don't have meaning at all, [that] they are not a part of language in some sense. But still, when he talks about how we determine what the reference of the name was, then he gives this picture. Unfortunately I don't have the passage in question with me, but this is what he says.⁹

The difference between using this theory as a theory of meaning and using it as a theory of reference will come out a little more clearly later on. But some of the attractiveness of the theory is lost if it isn't supposed to give the meaning of the name; for some of the solutions of problems that I've just mentioned will not be right, or at least won't clearly be right, if the description doesn't give the meaning of the name. For example, if someone said 'Aristotle does not exist' means 'there is no man doing such and such', or in the example from Wittgenstein, 'Moses does not exist', means 'no man did such and such', that might depend (and in fact, I think, does depend) on taking the theory in question as a theory of the meaning of the name 'Moses', not just as a theory of its reference. Well, I don't know. Perhaps all that is immediate now is the other way around: if 'Moses' means the same as 'the man who did such and such' then to say that Moses did not exist is to say that the man who did such and such did not exist, that is, that no one person did such and such. If, on the other hand, 'Moses' is not synonymous with any description, then even if its reference is in some sense determined by a description, statements containing the name cannot in general be analyzed by replacing the name by a description, though they may be materially equivalent to statements containing a description. So the analysis of singular existence statements mentioned above will have to be given up, unless it is established by some special argument, independent of a general

a r
all
for
used
had
prob
of id
and
a r
b w
all
and
of
pla

theory of the meaning of names; and the same applies to identity statements. In any case, I think it's false that 'Moses exists' means that at all. So we won't have to see if such a special argument can be drawn up.¹⁰

Before I go any further into this problem, I want to talk about another distinction which will be important in the methodology of these talks. Philosophers have talked (and, of course, there has been considerable controversy in recent years over the meaningfulness of these notions) [about] various categories of truth, which are called '*a priori*', 'analytic', 'necessary', – and sometimes even 'certain' is thrown into this batch. The terms are often used as if *whether* there are things answering to these concepts is an interesting question, but we might as well regard them all as meaning the same thing. Now, everyone remembers Kant (a bit) as making a distinction between '*a priori*' and 'analytic'. So maybe this distinction is still made. In contemporary discussion very few people, if any, distinguish between the concepts of statements being *a priori* and their being necessary. At any rate I shall *not* use the terms '*a priori*' and 'necessary' interchangeably here.

Consider what the traditional characterizations of such terms as '*a priori*' and 'necessary' are. First the notion of a *prioricity* is a concept of epistemology. I guess the traditional characterization from Kant goes something like: *a priori* truths are those which can be known independently of any experience. This introduces another problem before we get off the ground, because there's another modality in the characterization of '*a priori*', namely, it is supposed to be something which *can* be known independently of any experience. That means that in some sense it's *possible* (whether we do or do not in fact know it independently of any experience) to know this independently of any experience. And possible for whom? For God? For the Martians? Or just for people with minds like ours? To make this all clear might [involve] a host of problems all of its own about what sort of possibility is in question here. It might be best therefore, instead of using the phrase '*a priori* truth', to the extent that one uses it at all, to stick to the question of whether a particular person or knower knows something *a priori* or believes it true on the basis of *a priori* evidence.

I won't go further too much into the problems that might arise with the notion of a *prioricity* here. I will say that some philosophers somehow change the modality in this characterization from *can* to *must*.

are
 distinguished
 notion

They think that if something belongs to the realm of *a priori* knowledge, it couldn't possibly be known empirically. This is just a mistake. Something may belong in the realm of such statements that can be known *a priori* but still may be known by particular people on the basis of experience. To give a really common sense example: anyone who has worked with a computing machine knows that the computing machine may give an answer to whether such and such a number is prime. No one has calculated or proved that the number is prime; but the machine has given the answer: this number is prime. We, then, if we believe that the number is prime, believe it on the basis of our knowledge of the laws of physics, the construction of the machine, and so on. We therefore do not believe this on the basis of purely *a priori* evidence. We believe it (if anything is *a posteriori* at all) on the basis of *a posteriori* evidence. Nevertheless, maybe this could be known *a priori* by someone who made the requisite calculations. So 'can be known *a priori*' doesn't mean 'must be known *a priori*'.

The second concept which is in question is that of necessity. Sometimes this is used in an epistemological way and might then just mean *a priori*. And of course, sometimes it is used in a physical way when people distinguish between physical and logical necessity. But what I am concerned with here is a notion which is not a notion of epistemology but of metaphysics, in some (I hope) nonpejorative sense. We ask whether something might have been true, or might have been false. Well, if something is false, it's obviously not necessarily true. If it is true, might it have been otherwise? Is it possible that, in this respect, the world should have been different from the way it is? If the answer is 'no', then this fact about the world is a necessary one. If the answer is 'yes', then this fact about the world is a contingent one. This in and of itself has nothing to do with anyone's knowledge of anything. It's certainly a philosophical thesis, and not a matter of obvious definitional equivalence, either that everything *a priori* is necessary or that everything necessary is *a priori*. Both concepts may be vague. That may be another problem. But at any rate they are dealing with two different domains, two different areas, the epistemological and the metaphysical. Consider, say, Fermat's last theorem - or the Goldbach conjecture. The Goldbach conjecture says that an even number greater than 2 must be the sum of two prime numbers. If this is true, it is presumably necessary, and, if it is false, presumably

necessarily false. We are taking the classical view of mathematics here and assume that in mathematical reality it is either true or false.

If the Goldbach conjecture is false, then there is an even number, n , greater than 2, and for no primes p_1 and p_2 , both $< n$, does $n = p_1 + p_2$. This fact about n , if true, is verifiable by direct computation, and thus is necessary if the results of arithmetical computations are necessary. On the other hand, if the conjecture is true, then every even number exceeding 2 is the sum of two primes. Could it then be the case that, although in fact every even number is the sum of two primes, there might have been an even number which was not the sum of two primes? What would that mean? Such a number would have to be one of 4, 6, 8, 10, ...; and, by hypothesis, since we are assuming Goldbach's conjecture to be true, each of these can be shown, by direct computation, to be the sum of two primes. Goldbach's conjecture, then, cannot be contingently true or false; whatever truth-value it has belongs to it by necessity.

But what we can say of course is that right now as far as we know, the question can come out either way. So, in the absence of a mathematical proof deciding this question, none of us has any *a priori* knowledge about this question in either direction. We don't know whether Goldbach's conjecture is true or false. So right now we certainly don't know anything *a priori* about it.

Perhaps it will be alleged that we *can* in principle know *a priori* whether it is true. Well, maybe we can. Of course an infinite mind which can search through all the numbers can or could. But I don't know whether a finite mind can or could. Maybe there just is no mathematical proof whatsoever which decides the conjecture. At any rate this might or might not be the case. Maybe there is a mathematical proof deciding this question; maybe every mathematical question is decidable by an intuitive proof or disproof. Hilbert thought so; others have thought not; still others have thought the question unintelligible unless the notion of intuitive proof is replaced by that of formal proof in a single system. Certainly no one formal system decides all mathematical questions, as we know from Gödel. At any rate, and this is the important thing, the question is not trivial; even though someone said that it's necessary, if true at all, that every even number is the sum of two primes, it doesn't follow that anyone knows anything *a priori* about it. It doesn't even seem to me to follow without some further philosophical argument (it is an interesting

philosophical question) that anyone *could* know anything *a priori* about it. The 'could', as I said, involves some other modality. We mean that even if no one, perhaps even in the future, knows or will know *a priori* whether Goldbach's conjecture is right, in principle there is a way, which *could* have been used, of answering the question *a priori*. This assertion is not trivial.

The terms 'necessary' and '*a priori*', then, as applied to statements are *not* obvious synonyms. There may be a philosophical argument connecting them, perhaps even identifying them; but an argument is required, not simply the observation that the two terms are clearly interchangeable. (I will argue below that in fact they are not even co-extensive – that necessary *a posteriori* truths, and probably contingent *a priori* truths, both exist.)

I think people have thought that these two things must mean the same for these reasons. First, if something not only happens to be true in the actual world but is also true in all possible worlds, then, of course, just by running through all the possible worlds in our heads, we ought to be able with enough effort to see, if a statement is necessary, that it is necessary, and thus know it *a priori*. But really this is not so obviously feasible at all.

Secondly, I guess it's thought that, conversely, if something is known *a priori* it must be necessary, because it was known without looking at the world. If it depended on some contingent feature of the actual world, how could you know it without looking? Maybe the actual world is one of the possible worlds in which it would have been false. This depends on the thesis that there can't be a way of knowing about the actual world without looking which wouldn't be a way of knowing the same thing about every possible world. This involves problems of epistemology and the nature of knowledge; and of course it is very vague as stated. But it is not really *trivial* either. More important than any particular example of something which is alleged to be necessary and not *a priori* or *a priori* and not necessary, is to see that the notions are different, that it's not trivial to argue on the basis of something's being something which maybe we can only know *a posteriori*, that it's not a necessary truth. It's not trivial just because something is known in some sense *a priori*, that what is known is a necessary truth.

Another term used in philosophy is 'analytic'. Here it won't be too

important to get any clearer about this in this talk. The common examples of analytic statements, nowadays, are like 'bachelors are unmarried'. Kant, (someone just pointed out to me) gives as an example 'gold is a yellow metal', which seems to me an extraordinary one, because it's something I think that can turn out to be false. At any rate, let's just make it a matter of stipulation that an analytic statement is in some sense true by virtue of its meaning and true in all possible worlds by virtue of its meaning. Then something which is analytically true will be both necessary and *a priori*. (That's sort of stipulative.)

Another category I mentioned was that of certainty. Whatever certainty is, it's clearly not obviously the case that everything which is necessary is certain. Certainty is another epistemological notion. Something can be known, or at least rationally believed, *a priori*, without being quite certain. You've read a proof in the math book; and, though you think it's correct, maybe you've made a mistake. You often do make mistakes of this kind. You've made a computation, perhaps with an error.

There is one more question I want to go into in a preliminary way. Some philosophers have distinguished between essentialism, the belief in modality *de re*, and a mere advocacy of necessity, the belief in modality *de dicto*. Now, some people say: Let's give you the concept of necessity.¹¹ A much worse thing, something creating great additional problems, is whether we can say of any particular that it has necessary or contingent properties, even make the distinction between necessary and contingent properties. Look, it's only a *statement* or a *state of affairs* which can be either necessary or contingent. Whether a *particular* necessarily or contingently has a certain property depends on the way it's described. This is perhaps closely related to the view that the way we refer to particular things is by a description. What is Quine's famous example? If we consider the number 9, does it have the property of necessary oddness? Has that number got to be odd in all possible worlds? Certainly it's true in all possible worlds, let's say, it couldn't have been otherwise, that *nine* is odd. Of course, 9 could also be equally well picked out as *the number of planets*. It is not necessary, not true in all possible worlds, that the number of planets is odd. For example if there had been eight planets, the number of planets would not have been odd. And so it's thought: Was it necessary or contingent that Nixon won the election? (It might seem contingent, unless one has some view of some inexorable processes....) But this is a

contingent property of Nixon only relative to our referring to him as 'Nixon' (assuming 'Nixon' doesn't mean 'the man who won the election at such and such a time'). But if we designate Nixon as 'the man who won the election in 1968', then it will be a necessary truth, of course, that the man who won the election in 1968, won the election in 1968. Similarly whether an object has the same property in all possible worlds depends not just on the object itself but on how it is described. So it's argued.

It is even suggested in the literature, that though a notion of necessity may have some sort of intuition behind it (we do think some things could have been otherwise; other things we don't think could have been otherwise), this notion [of a distinction between necessary and contingent properties] is just a doctrine made up by some bad philosopher, who (I guess) didn't realize that there are several ways of referring to the same thing. I don't know if some philosophers have not realized this; but at any rate it is very far from being true that this idea [that a property can be held to be essential or accidental to an object independently of its description] is a notion which has no intuitive content, which means nothing to the ordinary man. Suppose that someone said, pointing to Nixon, 'That's the guy who might have lost'. Someone else says 'Oh no, if you describe him as 'Nixon', then he might have lost; but, of course, describing him as the winner, then it is not true that he might have lost'. Now which one is being the philosopher, here, the unintuitive man? It seems to me obviously to be the second. The second man has a philosophical theory. The first man would say, and with great conviction, "Well, of course, the winner of the election *might have been someone else*. The actual winner, had the course of the campaign been different, might have been the loser, and someone else the winner; or there might have been no election at all. So, such terms as 'the winner' and 'the loser' don't designate the same objects in all possible worlds. On the other hand, the term 'Nixon' is just a *name of this man*". When you ask whether it is necessary or contingent that *Nixon* won the election, you are asking the intuitive question whether in some counterfactual situation, *this man* would in fact have lost the election. If someone thinks that the notion of a necessary property (forget whether there *are* any necessary properties [and consider] just the *meaningfulness* of the notion.¹²) is a philosopher's notion with no intuitive content, he is wrong. Of course, some philosophers think that something's having intuitive content is very inconclusive

evidence in favor of it. I think it is very heavy evidence in favor of anything, myself. I really don't know in a way what more conclusive evidence one can have about anything, ultimately speaking. But, in any event, people who think the notion of accidental property unintuitive have intuition reversed, I think.

Why have they thought this? While there are many motivations for people thinking this, one is this: The question of essential properties so-called is supposed to be equivalent (and it is equivalent) to the question of identity across possible worlds. Suppose we have someone, Nixon, and there's another possible world where there is no one with all the properties Nixon has in the actual world. Which one of these other people, if any, is Nixon? Surely you must give some criterion of identity here. If you have a criterion of identity, then you just look in the other possible worlds at the man who is Nixon and the question whether in that other possible world Nixon has certain properties is well defined. It is also supposed to be well defined, in terms of such notions, whether it's true in all possible worlds, or there are some possible worlds in which Nixon didn't win the election. But, it's said, the problems of giving such criteria of identity are very difficult. Sometimes in the case of numbers it might seem easier (but even here it's argued that it's quite arbitrary). For example, one might say, and this is surely the truth, that if position in the series of numbers is what makes the number 9 what it is, then if (in another world) the number of planets had been 8, the number of planets would be a different number from the one it actually is. You wouldn't say that that number then is to be identified with our number 9 in this world. In the case of other types of objects, say people, material objects, things like that, has anyone given a set of necessary and sufficient conditions for identity across possible worlds?

Really, adequate necessary and sufficient conditions for identity which do not beg the question are very rare in any case. Mathematics is the only case I really know of where they are given even *within* a possible world, to tell the truth. I don't know of such conditions for identity of material objects over time, or for people. Everyone knows what a problem this is. But, let's forget about that. What seems to be more objectionable is that this depends on the wrong way of looking at what a possible world is. One thinks, in this picture, a possible world as if it were like a foreign country. One looks upon it as an observer. May-

be Nixon has moved to the other country and maybe he hasn't, but one is given only qualities. One can observe all his qualities, but of course, one doesn't observe that someone is Nixon. One observes that something has red hair (or green or yellow) but not whether something is Nixon. So we had better have a way of telling in terms of properties when we run into the same thing again as we saw before; we had better have a way of telling, when we come across one of these other possible worlds, who was Nixon.

Some logicians in their formal treatment of modal logic may encourage this picture. A prominent example, perhaps, is myself. Nevertheless, intuitively speaking, it seems to me not to be the right way of thinking about the possible worlds. A possible world isn't a distant country that we are coming across, or viewing through a telescope. Generally speaking, another possible world is too far away. Even if we travel faster than light, we won't get to it. A possible world is given by the descriptive conditions we associate with it. What do we mean when we say 'In some other possible worlds I might not have given this lecture today'? We just imagine the situation where I didn't decide to give this lecture or decided to give it on some other day. Of course, we don't imagine everything that is true or false, but only those things relevant to my giving the lecture; but, in theory, everything needs to be decided to make a total description of the world. We can't really imagine that except in part; that, then, is a 'possible world'. Why can't it be part of the *description* of a possible world that it contains *Nixon* and that in that world *Nixon* didn't win the election? It might be a question, of course, whether such a world *is* possible. (Here it would seem, *prima facie*, to be clearly possible.) But, once we see that such a situation is possible, then we are given that the man who might have lost the election or did lose the election in this possible world is Nixon, because that's part of the description of the world. 'Possible worlds' are stipulated, not discovered by powerful telescopes. There is no reason why we cannot stipulate that, in talking about what would have happened to Nixon in a certain counterfactual situation, we are talking about what would have happened to *him*.

Of course, if someone makes the demand that every possible world has to be described in a purely qualitative way, we can't say, 'Suppose Nixon had lost the election', we could say instead 'suppose a man with a dog named Checkers, who looks like a certain David Fry impersonation,

idea
had
via
part
of the
world

We'd
come
to
w

part
description
of
the
world

all
not
discovered

by
powerful
telescopes

is in a certain possible world and loses the election', Well, does he resemble Nixon enough to be identified with Nixon? A very explicit and blatant example of this way of looking at things is David Lewis' counterpart theory,¹³ but the literature on quantified modality is replete with it.¹⁴ Why need we make this demand? That is not the way we ordinarily think of counterfactual situations. We just say 'suppose this man had lost'. It is *given* that the possible world contains *this man*, and that in that world, he had lost. There may be a problem about what intuitions about possibility come to. But, if we have such an intuition about the possibility of *that* (*this man's* electoral loss), then it is about the possibility of *that*. It need not be identified with the possibility of a man looking like such and such, or holding such and such political views, or otherwise qualitatively described, having lost. We can point to the *man*, and ask what might have happened to *him*, had events been different.

It might be said 'Let's suppose that this is true. It comes down to the same thing, because whether Nixon could have had certain properties, different from the ones he actually has, is equivalent to the question of whether the criteria of identity across possible worlds include that Nixon does not have these properties'. But it doesn't really come to the same thing, because the usual notion of a criterion of transworld identity demands that we give purely qualitative necessary and sufficient conditions for someone being Nixon. If we can't imagine a possible world in which Nixon doesn't have a certain property, then it's a necessary condition of someone being Nixon. Or a necessary property of Nixon that he [has] that property. For example, supposing Nixon is in fact a human being, we might not imagine that there could have been a possible world in which he was, say, an inanimate object; perhaps it is not even possible for him not to have been a human being. Then it will be a necessary fact about Nixon that in all possible worlds where he exists at all, he is human or anyway he is not an inanimate object. This has nothing to do with any requirement that there be purely qualitative *sufficient* conditions for Nixonhood which we can spell out. And should there be? Maybe there is some argument that there should be, but we can consider these questions about *necessary* conditions without going into any question about *sufficient* conditions. Further, even if there were a purely qualitative set of necessary and sufficient conditions for being Nixon, the view I advocate would not demand that we find these conditions before we

can ask whether Nixon might have won the election, nor does it demand that we restate the question in terms of such conditions. We can simply consider Nixon and ask what might have happened to him had various circumstances been different. So the two views, the two ways of looking at things, do seem to me to make a difference.

Notice this question, whether Nixon could not have been a human being, is a clear case where the question asked is not epistemological. Suppose Nixon actually turned out to be an automaton. That might happen. We might need evidence for whether Nixon is a human being or an automaton. But that is a question about our knowledge. The question of whether Nixon might have not been a human being, given that he is one, is not a question about knowledge, *a posteriori* or *a priori*. It's a question about, even though such and such things are the case, what might have been the case otherwise.

This table is composed of molecules. Might it not have been composed of molecules? Certainly it was a scientific discovery of great moment that it was composed of molecules (or atoms). But could anything be this very object and not be composed of molecules? Certainly there is some feeling that the answer to that must be 'no'. At any rate it's hard to imagine under what circumstances you would have this very object and find that it is not composed of molecules. A quite different question is whether it is in fact composed of molecules in the actual world and how we know this. (I will go into more detail about these questions about essence later on.)

I wish at this point to introduce something which I need in the methodology of discussing the theory of names that I'm talking about. We need the notion of 'identity across possible worlds' as it's usually and, as I think, somewhat misleadingly called,¹⁵ to explicate one distinction that I want to make now. What's the difference between asking whether it's necessary that 9 is greater than 7 or whether it's necessary that the number of planets be greater than 7? Why does one show anything more about essence than the other? The answer to this might be intuitively 'Well, look, the number of planets might have been different from what it in fact is. It doesn't make any sense, though, to say that nine might have been different from what it in fact is'. Let's use some terms quasi-technically. Let's call something a *rigid designator* if in any possible world it designates the same object, a *non rigid* or *accidental designator* if that is

SAUL A. KRIPKE
 purely qualitative desc. of ~~world~~
 or counterfactual situation

not the case. Of course we don't require that the objects exist in all possible worlds. Certainly Nixon might not have existed if his parents had not gotten married, in the normal course of things. When we think of a property as essential to an object we usually mean that it is true of that object in any case where it would have existed. A rigid designator of a necessary existent can be called *strongly rigid*.

One of the intuitive theses I will maintain in these talks is that *names* are rigid designators. Certainly they seem to satisfy the intuitive test mentioned above: although someone other than the U.S. President in 1970 might have been the U.S. President in 1970 (e.g., Humphrey might have), no one other than Nixon might have been Nixon. In the same way, a designator rigidly designates a certain object if it designates that object wherever the object exists; if, in addition, the object is a necessary existent, the designator can be called *strongly rigid*. For example, 'the President of the U.S. in 1970' designates a certain man, Nixon; but someone else (e.g., Humphrey) might have been the President in 1970, and Nixon might not have; so this designator is not rigid.

In these lectures, I will argue, intuitively, that proper names are rigid designators, for ~~although the man (Nixon) might not have been the President, it is not the case that he might not have been Nixon (though he might not have been called 'Nixon')~~. Those who have argued that to make sense of the notion of rigid designator, we must antecedently make sense of 'criteria of transworld identity' have precisely reversed the cart and the horse: it is because we can refer (rigidly) to Nixon, and stipulate that we are speaking of what might have happened to him (under certain circumstances), that 'transworld identifications' are unproblematic in such cases.¹⁶

The tendency to demand purely qualitative descriptions of counterfactual situations has many sources. One, perhaps, is the confusion of the epistemological and the metaphysical, between a prioricity and necessity. If someone identifies necessity with a prioricity, and thinks that objects are named by means of uniquely identifying properties, he may think that it is the properties used to identify the object which, being known about it *a priori*, must be used to identify it in all possible worlds, to find out which object is Nixon. As against this, I repeat: (1) Generally, things aren't 'found out' about a counterfactual situation, they are stipulated; (2) possible worlds need not be given purely qualitatively, as

Argue
 must
 of no

if we were looking at them through a telescope. And we will see shortly that the properties an object has in a counterfactual world have nothing to do with the properties used to identify it in the actual world.¹⁷

Does the 'problem' of 'transworld identification' make any sense? Is it *simply* a pseudo-problem? The following, it seems to me, can be said for it. Although the statement that England fought Germany in 1943 perhaps cannot be *reduced* to any statement about individuals, nevertheless in some sense it is not a fact 'over and above' the collection of all facts about persons, and their behavior over history. The sense in which facts about nations are not facts 'over and above' those about persons can be expressed in the observation that a description of the world mentioning all facts about persons but omitting those about nations can be a complete description of the world, from which the facts about nations follow. Similarly, perhaps, facts about material objects are not facts 'over and above' facts about their constituent molecules. We may then ask, given a description of a non-actualized possible situation in terms of people, whether England still exists in that situation, or whether a certain nation (described, say, as the one where Jones lives) which would exist in that situation, is England. Similarly, given certain counterfactual vicissitudes in the history of the molecules of a table, *T*, one may ask whether *T* would exist, in that situation, or whether a certain bunch of molecules, which in that situation would constitute a table, constitute the very same table *T*. In each case, we ask criteria of identity across possible worlds for certain particulars in terms of those for other, more 'basic', particulars. If statements about nations (or tribes) are not *reducible* to those about other more 'basic' constituents, if there is some 'open texture' in the relationship between them, we can hardly expect to give hard and fast identity criteria; nevertheless, in concrete cases we may be able to answer whether a certain bunch of molecules would still constitute *T*, though in some cases the answer may be indeterminate. I think similar remarks apply to the problem of identity over time; here too we are usually concerned with determinacy, the identity of a 'complex' particular in terms of more 'basic' ones. (For example, if various parts of a table are replaced, is it the same object?¹⁸)

Such conception of 'transworld identification', however, differs considerably from the usual one. First, although we can try to describe the world in terms of molecules, there is no impropriety in describing it in

Bull Taurinus / Taurus

terms of grosser entities: the statement that *this table* might have been placed in another room is perfectly proper, in and of itself. We *need* not use the description in terms of molecules, or even grosser parts of the table, though we *may*. Unless we assume that some particulars are 'ultimate', 'basic' particulars, no type of description need be regarded as privileged. We can ask whether *Nixon* might have lost the election without further subtlety, and usually no further subtlety is required. Second, it is not assumed that necessary and sufficient conditions for what kinds of collections of molecules make up this table are possible; this fact I just mentioned. Third, the attempted notion deals with criteria of identity of particulars in terms of other *particulars*, not qualities. I can refer to the table before me, and ask what might have happened to it under certain circumstances; I can also refer to its molecules. If, on the other hand, it is demanded that I describe each counterfactual situation purely qualitatively, then I can only ask whether *a table*, of such and such color, and so on, would have certain properties; whether the table in question would be *this table*, table *T*, is indeed moot, since all reference to objects, as opposed to qualities, has disappeared. It is often said that, if a counterfactual situation is described as one which would have happened to *Nixon*, and if it is not assumed that such a description is reducible to a purely qualitative one, then mysterious 'bare particulars' are assumed, propertyless substrata underlying the qualities. This is not so: I think that *Nixon* is a Republican, not merely that he lies in back of Republicanism, whatever that means; I also think he might have been a Democrat. The same holds for any other properties *Nixon* may possess, except that some of these properties may be essential. What I do deny is that a particular is nothing but a 'bundle of qualities', whatever that may mean. If a quality is an abstract object, a bundle of qualities is an object of an even higher degree of abstraction, not a particular. Philosophers have come to the opposite view through a false dilemma: they have asked, are these objects *behind* the bundle of qualities, or is the object *nothing but* the bundle? Neither is the case; this table is wooden, brown, in the room, etc. It has all these properties and is not a thing without properties, behind them; but it should not therefore be identified with the set, or 'bundle', of its properties, nor with the subset of its essential properties. Don't ask: how can I identify this table in another possible world, except by its properties? I have the table in my hands,

by stipulation

I can point to it, and when I ask whether *it* might have been in another room, I am talking, by definition, about *it*. I don't have to identify it after seeing it through a telescope. If I am talking about it, I am talking about *it*, in the same way as when I say that our hands might have been painted green, I have stipulated that I am talking about greenness. Some properties of an object may be essential to it, in that it could not have failed to have them; but these properties are not used to identify the object in another possible world, for such an identification is not needed; nor need the essential properties of an object be the properties used to identify it in the actual world, if indeed it is identified in the actual world by means of properties (I have up to now left the question open).

So: the question of transworld identification makes *some* sense, in terms of asking about the identity of an object *via* questions about its component parts. But these parts are not qualities, and it is not an object resembling the given one which is in question. Theorists have often said that we identify objects across possible worlds as objects resembling the given one in the most important respects. On the contrary, Nixon, had he decided to act otherwise, might have avoided politics like the plague, through privately harboring radical opinions. Most important, even when we can replace questions about an object by questions about its parts, we need not do so. We can refer to the object and ask what might have happened to *it*. So, we do not begin with worlds (which are supposed somehow to be real, and whose qualities, but not whose objects, are perceptible to us), and then ask about criteria of transworld identification; on the contrary, we begin with the objects, which we *have*, and can identify, in the actual world. We can then ask whether certain things might have been true of the objects.

I mentioned the distinction between this question about the theory of naming as the theory of meaning or as a theory of reference. Let me give an example of this. Suppose someone stipulates that 100 degrees centigrade is to be the temperature at which water boils at sea level. This isn't completely precise because the pressure may vary at sea level. Of course, historically, a more precise definition was given later. But let's suppose that this were the definition. Another sort of example in the literature is that one meter is to be the length of *S* where *S* is a certain stick or bar in Paris. (Usually people who like to talk about these de-

initions then try to make 'the length of' into an 'operational' concept. But it's not important.)

Wittgenstein says something very puzzling about this. He says: "There is one thing of which one can say neither that it is one meter long nor that it is not one meter long, and that is the standard meter in Paris. But this is, of course, not to ascribe any extraordinary property to it, but only to mark its peculiar role in the language game of measuring with a meter rule."¹⁹ This seems to be a very 'extraordinary property', actually, for any stick to have. I think he must be wrong. If the stick is a stick, for example, 39.37 inches long (I assume we have some different standard for inches), why isn't it one meter long? Anyway, let's suppose that he is wrong and that the stick is one meter long. Part of the problem which is bothering Wittgenstein is, of course, that this stick serves as a standard of length and so we can't attribute length to it. Be this as it may (well, it may not be), is the statement 'stick *S* is one meter long', a necessary truth? Of course its length might vary in time. We could make the definition more precise by stipulating that one meter is to be the length of *S* at a fixed time t_0 . Is it then a necessary truth that stick *S* is one meter long at time t_0 ? Someone who thinks that everything one knows *a priori* is necessary might think: "This is the *definition* of a meter. By definition, stick *S* is one meter long at t_0 . That's a necessary truth." But there seems to me to be no reason so to conclude, even for a man who uses the stated definition of 'one meter'. For he's using this definition not to give the meaning of what he called the 'meter', but to fix the reference. (For such an abstract thing as a unit of length, the notion of reference may be unclear. But let's suppose it's clear enough for the present purposes.) He uses it to fix a reference. There is a certain length which he wants to mark out. He marks it out by an accidental property, namely that there is a stick of that length. Someone else might mark out the same reference by another accidental property. But in any case, even though he uses this to fix the reference of his standard of length, a meter, he can still say, "if heat had been applied to this stick *S* at t_0 , then at t_0 stick *S* would not have been one meter long."

Well, why can he do this? Part of the reason may lie in some people's minds in the philosophy of science, which I don't want to go into here. But a simple answer to the question is this: Even if this is the *only* standard of length that he uses,²⁰ there is an intuitive difference between the

Definition

phrase 'one meter' and the phrase 'the length of S at t_0 '. The first phrase is meant to designate rigidly a certain length in all possible worlds, which in the actual world happens to be the length of the stick S at t_0 . On the other hand 'the length of S at t_0 ' does not designate anything rigidly. In some counterfactual situations the stick might have been longer and in some shorter if various stresses and strains had been applied to it. So we can say of this stick the same way as we would of any other of the same substance and length, that if heat of a given quantity had been applied to it, it would have expanded to such and such a length. Such a counterfactual statement, being true of other sticks with identical physical properties, will also be true of this stick. There is no conflict between that counterfactual statement and the definition of 'one meter' as 'the length of S at t_0 ', because the 'definition', properly interpreted, does not say that the phrase 'one meter' is to be *synonymous* (even when talking about counterfactual situations) with the phrase 'the length of S at t_0 ', but rather that we have *determined the reference* of the phrase 'one meter' by stipulating that 'one meter' is to be a *rigid* designator of the length which is in fact the length of S at t_0 . So this does not make it a necessary truth that S is one meter long at t_0 . In fact, under certain circumstances, S would not have been one meter long. The reason is that one designator ('one meter') is rigid and the other designator ('the length of S at t_0 ') is not.

What then, is the *epistemological* status of the statement 'Stick S is one meter long at t_0 ', for someone who has fixed the metric system by reference to stick S ? It would seem that he knows it *a priori*. For if he used stick S to fix the reference of the term 'one meter', then as a result of this kind of 'definition' (which is not an abbreviative or synonymous definition), he knows automatically, without further investigation, that S is one meter long.²¹ On the other hand, even if S is used as the standard of a meter, the *metaphysical* status of ' S is one meter long' will be that of a contingent statement, provided that 'one meter' is regarded as a rigid designator: under appropriate stresses and strains, heatings or coolings, S would have had a length other than one meter even at t_0 . (Such statements as, 'Water boils at 100°C at sea level', can have a similar status.) So in this sense, there are contingent *a priori* truths. More important for present purposes, though, than accepting this example as an instance of the contingent *a priori*, is its illustration of the distinction between 'definitions' which fix a reference and those which give a synonymy.

2 kinds of defn.

one is rigid
length is not
non-
defining
does not
give a
synonym
it fix
refer
of defn
from

It cannot be rigid.

In the case of names one might make this distinction too. Suppose the reference of a name is given by a description or a cluster of descriptions. If the name *means the same* as that description or cluster of descriptions, it will not be a rigid designator. It will not necessarily designate the same object in all possible worlds, since other objects might have had the given properties in other possible worlds, unless (of course) we happened to use essential properties in our description. So suppose we say, 'Aristotle is the greatest man who studied under Plato'. If we used that as a *definition*, the name 'Aristotle' is to mean 'the greatest man who studied under Plato'. Then of course in some other possible world that man might not have studied under Plato and some other man would have been Aristotle. If, on the other hand, we merely use the description to *fix the referent* then that man will be the referent of 'Aristotle' in all possible worlds. The only use of the description will have been to pick out to which man we mean to refer. But then, when we say counterfactually 'suppose Aristotle had never gone into philosophy at all', we need not mean 'suppose a man who studied with Plato, and taught Alexander the Great, and wrote this and that, and so on, had never gone into philosophy at all', which might seem like a contradiction. We need only mean, 'suppose that *that man* had never gone into philosophy at all'.

It seems plausible to suppose that, in some cases, the reference of a name is indeed fixed *via* a description in the same way that the metric system was fixed.

When the mythical agent first saw Hesperus, he may well have fixed his reference by saying, "I shall use 'Hesperus' as a name of the heavenly body appearing in yonder position in the sky." He then has fixed the reference of 'Hesperus' by its apparent celestial position. Does it follow that it is part of the *meaning* of the name that Hesperus has such and such position at the time in question? Surely not: if Hesperus had been hit earlier by a comet, it might have been visible at a different position at that time. In such a counterfactual situation we would say that Hesperus would not have occupied that position, but not that Hesperus would not have been Hesperus. The reason is that 'Hesperus' rigidly designates a certain heavenly body and 'the body in yonder position' does not – a different body, or no body might have been in that position, but no other body might have been Hesperus (though another body, not Hesperus,

2 Over names meaning from ordinary use
names denotate from ordinary use

might have been called 'Hesperus'). Indeed, as I have said, I will hold that names are always rigid designators.

Frege and Russell certainly seem to have the full blown theory according to which a proper name is not a rigid designator and is synonymous with the description which replaced it. But another theory might be that this description is used to determine a rigid reference. These two alternatives will have different consequences for the questions I was asking before. If 'Moses' means 'the man who did such and such', then, if no one did such and such, Moses didn't exist; and maybe 'no one did such and such' is even an *analysis* of 'Moses didn't exist'. But if the description is used to fix a reference rigidly, then it's clear that that is *not* what is meant by 'Moses didn't exist', because we can ask, if we speak of a counterfactual case where no one did indeed do such and such, say, lead the Israelites out of Egypt, does it follow that, in such a situation, Moses wouldn't have existed? It would seem not. For surely Moses might have just decided to spend his days more pleasantly in the Egyptian courts. He might never have gone into either politics or religion at all; and in that case maybe no one would have done any of the things that the Bible relates of Moses. That doesn't in itself mean that in such a possible world Moses wouldn't have existed. If so, then 'Moses exists' means something different from 'the existence and uniqueness conditions for a certain description are fulfilled'; and therefore this does not give an analysis of the singular existential statements after all. If you give up the idea that this is a theory of meaning and make it into a theory of reference in the way that I have described it, you give up some of the advantages of the theory. Singular existential statements and identity statements between names need some other analysis.

Frege should be criticized for using the term 'sense' in two senses. For he takes the sense of a designator to be its meaning; and he also takes it the way its reference is determined. Identifying the two, he supposes that both are given by definite descriptions. Ultimately, I will reject this second supposition too; but even were it right, I reject the first. A description may be used as synonymous with a designator, or it may be used to fix its reference. The two Fregean senses of 'sense' correspond to two senses of 'definition' in ordinary parlance. They should carefully be distinguished.²²

I hope the idea of fixing the reference as opposed to actually defining

one term as meaning the other is somewhat clear. There is really not enough time to go into every thing in great detail. I think, even in cases where the notion of rigidity versus accidentality of designation cannot be used to make out the difference in question, some things called definitions really intend to fix a reference rather than to give the meaning of a phrase, to give a synonym. Let me give an example. π is supposed to be the ratio of the circumference of a circle to its diameter. Now, it's something that I have nothing but a vague intuitive feeling to argue for: It seems to me that here this Greek letter is not being used as short for the phrase 'the ratio of the circumference of a circle to its diameter' nor is it even used as short for a cluster of alternative definitions of π , whatever that might mean. It is used as a name for a real number, which in this case is necessarily the ratio of the circumference of a circle to its diameter. Note that here both ' π ' and the 'ratio of the circumference of a circle to its diameter' are rigid designators, so the arguments given in the metric case are inapplicable. (Well, if someone doesn't see this, or thinks it's wrong, it doesn't matter.)

Let me return to the question about names which I raised. As I said, there is a popular modern substitute for the theory of Frege and Russell; it is adopted even by such a strong critic of many views of Frege and Russell, especially the latter, as Strawson.²³ The substitute is that, although a name is not a disguised description it either abbreviates, or anyway its reference is determined by, some cluster of descriptions. The question is whether this is true. And as I say, there are stronger and weaker versions of this. The stronger version would say that the name is simply *defined*, synonymously, as the cluster of descriptions. It will then be necessary, not that Moses had any particular property in this cluster, but that he had the disjunction of them. There couldn't be any counterfactual situation in which he didn't do any of those things. I think it's clear that I think this is very implausible. People *have* said it – or maybe they haven't been intending to say that, but were using 'necessary' in some other sense. At any rate, for example, in Searle's article on proper names:

To put the same point differently, suppose we ask, 'why do we have proper names at all?' Obviously to refer to individuals. 'Yes but descriptions could do that for us'. But only at the cost of specifying identity conditions every time reference is made: Suppose we agree to drop 'Aristotle' and use, say, 'the teacher of Alexander', then it is a necessary truth that the man referred to is Alexander's teacher – but it is a contingent fact that Aristotle ever went into pedagogy (though I am suggesting that it

is a necessary fact that Aristotle has the logical sum, inclusive disjunction, of properties commonly attributed to him).²⁴

Such a suggestion, if 'necessary' is used in the way I have been using it in this lecture, must clearly be false. (Unless he's got some very interesting essential property commonly attributed to Aristotle.) Most of the things commonly attributed to Aristotle are things that Aristotle might not have done at all. In a situation in which he didn't do them, we would describe that as a situation in which *Aristotle* didn't do them. This is not a distinction of scope, as happens sometimes in the case of descriptions, where someone might say that the man who taught Alexander might not have taught Alexander; though it could not have been true that: the man who taught Alexander didn't teach Alexander. This is Russell's distinction of scope. (I won't go into it.) It seems to me clear that this is not the case here. Not only is it true of the man Aristotle that he might not have gone into pedagogy; it is also true that we use the term 'Aristotle' in such a way that, in thinking of a counterfactual situation in which Aristotle didn't go into any of the fields and do any of the achievements we commonly attribute to him, still we would say that was a situation in which *Aristotle* did not do these things.²⁵ Well there are some things like the date, the period he lived, that might be more imagined as necessary. Maybe those are things we commonly attribute to him. There are exceptions. Maybe it's hard to imagine how he could have lived 500 years later than he in fact did. That certainly raises at least a problem. But take a man who doesn't have any idea of the date. Many people just have some vague cluster of his most famous achievements. Not only each of these singly, but the possession of the entire disjunction of these properties, is just a contingent fact about Aristotle; and the statement that Aristotle had this disjunction of properties is a contingent truth.

A man might know it *a priori* in some sense, if he in fact fixes the reference of 'Aristotle' as the man who did one of these things. Still it won't be a necessary truth for him. So this sort of example would be an example where *a prioricity* would not necessarily imply necessity, if the cluster theory of names were right. The case of fixing the reference of 'one meter' is a very clear example in which someone, just because he fixed the reference in this way, can in some sense know *a priori* that the length of this stick is a meter without regarding it as a necessary truth. Maybe the thesis about *a prioricity* implying necessity can be modified.

It does appear to state some insight which might be important, and true, about epistemology. In a way an example like this may seem like a trivial counterexample which is not really the point of what some people think when they think that only necessary truths can be known *a priori*. Well, if the thesis that all *a priori* truth is necessary is to be immune from this sort of counterexample, it needs to be modified in some way. Unmodified it leads to confusion about the nature of reference. And I myself have no idea how it should be modified or restated, or if such a modification or restatement is possible.²⁶

Let me state then what the cluster concept theory of names is. (It really is a nice theory. The only defect I think it has is probably common to all philosophical theories. It's wrong. You may suspect me of proposing another theory in its place; but I hope not, because I'm sure it's wrong too if it is a theory.) The theory in question can be broken down into a large number of theses, especially if you want to see how it handles the problem of identity statements, existence statements, and so on. There are more statements if you take it in the stronger version as a theory of meaning. The speaker is *A*.

- (1) To every name or designating expression '*X*', there corresponds a cluster of properties, namely the family of properties ϕ such that *A* believes ' ϕX '.

This thesis is true, because it can just be a definition. Now, of course, some people might think that not everything the speaker believes about *X* has anything to do with determining the reference of '*X*'. They might only be interested in a subset. But we can handle this later on by modifying some of the other features. So this thesis is correct, by definition. The theses that follow, however, are all, I think, false.

- (2) One of the properties, or some conjointly, are believed by *A* to pick out some individual uniquely.

This doesn't say that they do pick out something uniquely, just that *A* believes that they do. Another thesis is that he is correct.

- (3) If most, or a weighted most, of the ϕ 's are satisfied by one unique object *y*, then *y* is the referent of '*X*'.

Well, the theory says that the referent of '*X*' is supposed to be the thing

satisfying, if not all the properties, 'enough' of them. Obviously *A* could be wrong about some things about *X*. You take some sort of a vote. Now the question is whether this vote should be democratic or have some inequalities among the properties. It seems more plausible that there should be some weighting, that some properties are more important than others. A theory really has to specify how this weighting goes. I believe that Strawson, to my surprise, explicitly states that democracy should rule here, so the most trivial properties are of equal weight with the most crucial.²⁷ Surely it is more plausible to suppose that there is some weighting. Let's say democracy doesn't necessarily rule. If there is any property that's completely irrelevant to the reference we can disenfranchise it altogether, by giving it weight 0. The properties can be regarded as members of a corporation. Some have more stock than others; some may even have only non-voting stock.

- (4) If the vote yields no unique object, '*X*' does not refer.
- (5) The statement, 'If *X* exists, then *X* has most of the ϕ 's' is known *a priori* by the speaker.
- (6) The statement, 'If *X* exists, then *X* has most of the ϕ 's' expresses a necessary truth (in the idiolect of the speaker).

(6) need not be a thesis of the theory if someone doesn't think that the cluster is part of the meaning of the name. He could think that though he determines the reference of Aristotle as the man who had most of the ϕ 's, still there are certainly possible situations in which Aristotle wouldn't have had most of the ϕ 's.

There are some subsidiary theses I was mentioning. These would give the analyses of singular existential statements, like 'Moses exists', means 'enough of the properties ϕ are satisfied'. Even the man who doesn't use the theory as a theory of meaning has some of these theses. For example, subsidiary to thesis 4, we should say that it is *a priori* true for the speaker that, if not enough of the ϕ 's are satisfied, then '*X*' does not exist. Only if he holds the view as a theory of meaning, rather than of reference, would it also be *necessarily* true that, if not enough of the ϕ 's are satisfied, '*X*' does not exist. In any case it will be something he knows *a priori*. (At least he will know it *a priori* provided he knows the proper theory of names.) Then there is also an analysis of identity statements along the same lines.

The question is, are any of these true? If true, they give a nice picture of what's going on. Preliminary to discussing these theses, let me mention that, often, when people specify which properties ϕ are relevant, they seem to specify them wrongly. That's just an incidental defect, though it is closely related to the arguments against the theory that I will give presently. Consider the example from Wittgenstein. What does he say the relevant properties are? "When one says 'Moses does not exist', this may mean various things. It may mean: the Israelites did not have a *single* leader when they withdrew from Egypt – or: their leader was not called Moses – or: there cannot have been anyone who accomplished all that the Bible relates of Moses" The gist of all this is that we know *a priori* that, if the Biblical story is substantially false, Moses did not exist. I have already argued that the Biblical story does not give *necessary* properties of Moses, that he might have lived without doing any of these things. Here I ask whether we know *a priori* that if Moses existed, he in fact did some or most of them. Is this really the cluster of properties that we should use here? Surely there is a distinction which is neglected in these kinds of remarks. The Biblical story might have been a complete legend, or it might have been a substantially false account of a real person. In the latter case, it seems to me that a scholar could say that he supposes that, though Moses did exist, the things said of him in the Bible are substantially false. Such things occur in this very field of scholarship. Suppose that someone says that no prophet ever was swallowed by a big fish or a whale. Does it follow, on that basis, that Jonah did not exist? There still seems to be the question whether the Biblical account is a legendary account of no person or a legendary account built on a real person. In the latter case, it's only natural to say that, though Jonah did exist, no one did the things commonly related to him. I choose this case because while Biblical scholars generally hold that Jonah did exist, the account not only of his being swallowed by a big fish but even going to Nineveh to preach or anything else that is said in the Biblical story is assumed to be substantially false. But nevertheless there are reasons for thinking this was about a real prophet. If I had a suitable book along with me I could start quoting out of it: 'Jonah, the Son of Amittai, was a real prophet, however such and such and such'. There are independent reasons for thinking this was not a pure legend about an imaginary character but one about a real character.²⁸

These examples could be modified. Maybe all we believe is that *the Bible relates of him* that such and such. This gives us another problem, because how do we know whom the Bible is referring to? The question of our reference is thrown back to the question of reference in the Bible. This leads to a condition which we ought to put in explicitly.

(C) For any successful theory, the account must not be circular. The properties which are used in the vote must not themselves involve the notion of reference in a way that it is ultimately impossible to eliminate.

Let me give an example where the noncircularity condition is clearly violated. The following theory of proper names is due to William Kneale in an article called 'Modality, De Dicto and De Re'.²⁹ It contains, I think, a clear violation of noncircularity conditions.

Ordinary proper names of people are not, as John Stuart Mill supposed, – signs without sense. While it may be informative to tell a man that the most famous Greek philosopher was called Socrates, it is obviously trifling to tell him that Socrates was called Socrates; and the reason is simply that he cannot understand your use of the word 'Socrates' at the beginning of your statement unless he already knows that 'Socrates' means 'The individual called 'Socrates''.³⁰

Here we have a theory of the reference of proper names. 'Socrates' just means 'the man called 'Socrates''. Actually, of course, maybe not just one man can be called 'Socrates', and some may call him 'Socrates' while others may not. Certainly that is a condition which under some circumstances is uniquely satisfied. Maybe only one man was called 'Socrates' by me on a certain occasion.

Kneale says it's trifling to tell someone that Socrates *was* called 'Socrates'. That isn't trifling on any view. Maybe the Greeks didn't call him 'Socrates'. Let's say that Socrates is called 'Socrates' by us – by *me* anyway. Suppose that's trifling. (I find it surprising that Kneale uses the past tense here; it is dubious that the Greeks *did* call him 'Socrates' – at least, the Greek name is pronounced differently. I will check the accuracy of the quotation for the next lecture.)

Kneale gives an argument for this theory. 'Socrates' must be analyzed as 'the individual called 'Socrates'', because how else can we explain the fact that it is trifling to be told that Socrates is called 'Socrates'? In some cases that's rather trifling. In the same sense, I suppose, you could get a good theory of the meaning of any expression in English and construct a dictionary. For example, though it may be informative to tell someone that horses are used in races, it is trifling to tell him that

horses are called 'horses'. Therefore this could only be the case because the term 'horse', means in English 'the things called 'horses' '. Similarly with any other expression which might be used in English. Since it's trifling to be told that sages are called 'sages', 'sages' just means 'the people called 'sages''. Now plainly this isn't really a very good argument, nor can it therefore be the only explanation of why it's trifling to be told that Socrates is called 'Socrates'. Let's not go into exactly why it's trifling. Of course, anyone who knows the use of 'is called' in English, even without knowing what the statement means, knows that if 'quarks' means something then 'quarks are called 'quarks'' will express a truth. He may not know what it expresses, because he doesn't know what a quark is. But that does not have much to do with the meaning of the term 'quarks'.

We could go into this actually at great length. There are interesting problems coming out of this sort of passage. But the main reason I wanted to introduce it here is that as a theory of reference it would give a clear violation of the noncircularity condition. Someone uses the name 'Socrates'. How are we supposed to know to whom he refers? By using the description which gives the sense of it. According to Kneale, the description is 'the man called 'Socrates''. And here, (presumably, since this is supposed to be so trifling!) it tells us nothing at all. Taking it in this way it seems to be no theory of reference at all. We ask to whom does he refer by 'Socrates' and then the answer is given as, well he refers to the man to whom he refers. If this were all there was to the meaning of a proper name, then no reference would get off the ground at all.

So there's a condition to be satisfied; in the case of this particular theory it's obviously unsatisfied. The paradigm amazingly enough, is even sometimes used by Russell as the descriptive sense, namely: 'the man called 'Walter Scott''. Obviously if the only descriptive senses of names we can think of are of the form 'the man called such and such', 'the man called 'Walter Scott'', 'the man called 'Socrates' ', then whatever this relation of *calling* is is really what determines the reference and not any description like 'the man called 'Socrates''.

LECTURE II: JANUARY 22, 1970

Last time we ended up talking about a theory of naming which is given by a number of theses here on the board.

- (1) To every name or designating expression ' X ', there corresponds a cluster of properties, namely the family of those properties ϕ such that A believes ' ϕX '.
- (2) One of the properties, or some conjointly, are believed by A to pick out some individual uniquely.
- (3) If most, or a weighted most, of the ϕ 's are satisfied by one unique object y , then y is the referent of ' X '.
- (4) If the vote yields no unique object, ' X ' does not refer.
- (5) The statement, 'If X exists, then X has most of the ϕ 's' is known *a priori* by the speaker.
- (6) The statement, 'If X exists, then X has most of the ϕ 's' expresses a necessary truth (in the idiolect of the speaker).
- (C) For any successful theory, the account must not be circular. The properties which are used in the vote must not themselves involve the notion of reference in such a way that it is ultimately impossible to eliminate.

(C) is not a thesis but a condition on the satisfaction of the other theses. In other words Theses (1)–(6) cannot be satisfied in a way which leads to a circle, in a way which does not lead to any independent determination of reference. The example I gave last time of a blatantly circular attempt to satisfy these conditions, was a theory of names mentioned by William Kneale. I was a little surprised at the statement of the theory when I was reading what I had copied down, so I looked it up again. I looked it up in the book to see if I'd copied it down accurately. Kneale *did* use the past tense. He said that though it is not trifling to be told that Socrates was the greatest philosopher of ancient Greece, it is trifling to be told that Socrates was called 'Socrates'. Therefore, he concludes, the name 'Socrates' must simply mean 'the individual called 'Socrates''. Russell, as I've said, in some places gives a similar analysis. Anyway as stated using the past tense, the condition wouldn't be circular, because one certainly could decide to use the term 'Socrates' to refer to whoever was called 'Socrates' by the Greeks. But, of course, in that sense it's not at all trifling to be told that Socrates was called 'Socrates'. If this is any kind of fact, it might be false. Perhaps we know that *we* call him 'Socrates'; that hardly shows that the Greeks did so. In fact, of course, they may have pronounced the name differently. It may be, in the case of this particular

name, that transliteration from the Greek is so good that the English version is not pronounced very differently from the Greek. But that won't be so in the general case. Certainly it is not trifling to be told that Isaiah was called 'Isaiah'. In fact, it is false to be told that Isaiah was called 'Isaiah'; the prophet wouldn't have recognized this name at all. And perhaps the Greeks didn't call their country anything like 'Greece'. Suppose we amend the thesis so that it reads: it's trifling to be told that Socrates is called 'Socrates' by us, or at least, by me, the speaker. Then in some sense this is fairly trifling. I don't think it is necessary or analytic. In the same way, it is trifling to be told that horses are called 'horses' without this leading to the conclusion that the word 'horse' simply means 'the animal called a 'horse''. As a theory of the referent of the name 'Socrates' it will lead immediately to a vicious circle. If one was determining the referent of a name like 'Glunk' to himself and made the following decision, 'I shall use the term 'Glunk' to refer to the man that I call 'Glunk'', this would get one nowhere. One had better have some independent determination of the referent of 'Glunk'. This is a good example of a blatantly circular determination. Actually sentences like 'Socrates is called 'Socrates'' are very interesting and one can spend, strange as it may seem, hours talking about their analysis. I actually did, once, do that. I won't do that, however, on this occasion. (See how high the seas of language can rise. And at the lowest points too.) Anyway this is a useful example of a violation of the noncircularity condition. The theory will satisfy all of these statements, perhaps, but it satisfies them only because there is some independent way of determining the reference independently of the particular condition: being the man called 'Socrates'.

I have already talked about, in the last lecture, Thesis (6). Theses (5) and (6), by the way, have converses. What I said for Thesis (5) is that the statement that if X exists, X has most of the ϕ 's is *a priori* true for the speaker. It will also be true under the given theory that certain converses of this statement hold true also *a priori* for the speaker, namely; if any unique thing has most of the properties ϕ in the properly weighted sense, it is X . Similarly a certain converse to this will be *necessarily* true, namely: if anything has most of the properties ϕ in the properly weighted sense, it is X . So really one can say that it is both *a priori* and necessary that something is X if and only if it uniquely has most of the properties ϕ . This really comes from the previous Theses (1)-(4), I suppose. And (5)

~ (The cluster of descriptions that pick out the referent of a name are necessary of that object)

and (6) really just say that a sufficiently reflective speaker grasps this theory of proper names. Knowing this, he therefore sees that (5) and (6) are true. The objections to Theses (5) and (6) will not be that some speakers are unaware of this theory and therefore don't know these things.

What I talked about in the last lecture is Thesis (6). It's been observed by many philosophers that, if the cluster of properties associated with a description is taken in a very narrow sense, so that only one property is given any weight at all, let's say one definite description to pick out the referent – for example, Aristotle was the philosopher who taught Alexander the Great – then certain things will seem to turn out to be necessary truths which are not necessary truths – in this case, for example, that Aristotle taught Alexander the Great. But as Searle said, it is not a necessary truth but a contingent one that Aristotle ever went into pedagogy. Therefore, he concludes that one must drop the original paradigm of a single description and turn to that of a cluster of descriptions.

To summarize some things that I argued last time, this is not the correct answer (whatever it may be) to this problem about necessity. For Searle goes on to say,

Suppose we agree to drop 'Aristotle' and use, say, 'the teacher of Alexander', then it is a necessary truth that the man referred to is Alexander's teacher – but it is a contingent fact that Aristotle ever went into pedagogy, though I am suggesting that it is a necessary fact that Aristotle has the logical sum, inclusive disjunction, of properties commonly attributed to him...³¹

~~This is what is not so. It just is not, in any intuitive sense of necessity, a necessary truth that Aristotle had the properties commonly attributed to him. There is a certain theory, perhaps popular in some views of the philosophy of history, which might both be deterministic and yet at the same time assign a great role to the individual in history. Perhaps Carlyle would associate with the meaning of the name of a great man his achievements. According to such a view it will be necessary, once a certain individual is born, that he is destined to perform various great tasks and so it will be part of the very nature of Aristotle that he should have produced ideas which had a great influence on the western world. Whatever the merits of such a view may be as a view of history or the nature of great men, it does not seem that it should be trivially true on the basis of a theory of proper names. It would seem that it's a contingent fact that~~

Aristotle ever did *any* of the things commonly attributed to him today, *any* of these great achievements that we so much admire. I must say that there is *something* to this feeling of Searle's. When I hear the name 'Hitler', I do feel it's sort of analytic that that man was evil. But really, probably not. Hitler might have spent all his days in quiet in Linz. In that case we would not say that then this man would not have been Hitler, for we use the name 'Hitler' just as the name of that man, even describing possible worlds. (This is the notion which I called a *rigid designator* in the previous talk.) Suppose we do decide to pick out the reference of 'Hitler', as the man who succeeded in having more Jews killed than anyone else managed to do in history. That is the way we pick out the reference of the name; but in another counterfactual situation where some one else would have gained this discredit, we wouldn't say that in that case that other man would have been Hitler. If Hitler had never come to power, Hitler would not have had the property which I am supposing we use to fix the reference of his name. Similarly, even if we define what a meter is by reference to the standard meter stick, it will be a contingent truth and not a necessary one that that particular stick is one meter long. If it had been stretched, it would have been longer than one meter. And that is because we use the term 'one meter' rigidly to designate a certain length. Even though we fix what length we are designating by an accidental property of that length, just as in the case of the name of the man we may pick the man out by an accidental property of the man, still we use the name to designate that man or that length in all possible worlds.

The property we use need not be one which is regarded in any way as necessary or essential. In the case of a yard, the original way this property was picked out was, I think, the distance when the arm of King Henry I of England was outstretched from the tip of his finger to his nose. If this was the length of a yard, it nevertheless will not be a necessary truth that the distance between the tip of his finger and his nose should be a yard. Maybe an accident might have happened to foreshorten his arm; that would be possible. And the reason that it's not a necessary truth is not that there might be other criteria in a 'cluster concept' of yardhood. Even a man who strictly uses King Henry's arm as his one standard of length can say, counterfactually, that if certain things had happened to the King, the exact distance between the end of one of his fingers and his nose would not have been exactly a yard. He need not be using a

Counterpart Theory Bank

Importance
289 *ess*

cluster as long as he uses the term 'yard' to pick out a certain fixed reference to be that length in all possible worlds.

These remarks show, I think, the intuitive bizarreness of a good deal of the literature on 'transworld identification' and 'counterpart theory'. For many theorists of these sorts, believing, as they do, that a 'possible world' is given to us only qualitatively, argue that Aristotle is to be 'identified in other possible worlds', or alternatively that his counterparts are to be identified, as those things in other possible worlds who most closely resemble Aristotle in his most important properties. (Lewis, for example, says: "Your counterparts ... resemble you ... in important respects ... more closely than do the other things in their worlds ... weighted by the importance of the various respects and by the degrees of the similarities."³²) Some may equate the important properties with those properties used to identify the object in the actual world.

Surely these notions are incorrect. To me Aristotle's most important properties consist in his philosophical work, and Hitler's in his murderous political role; both, as I have said, might have lacked these properties altogether. Surely there was no logical fate hanging over either Aristotle or Hitler which made it in any sense inevitable that they should have possessed the properties we regard as important to them; they could have had careers completely different from their actual ones. Important properties of an object need not be essential, unless 'importance' is used as a synonym for essence; and an object could have had properties very different from its most striking actual properties, or from the properties we use to identify it.

To clear up one thing which some people have asked me: When I say that a designator is rigid, and designates the same thing in all possible worlds, I mean that, as used in *our* language, it stands for that thing, when we talk about counterfactual situations. I don't mean, of course, that there mightn't be counterfactual situations in which in the other possible worlds people actually spoke a different language. One doesn't say that 'two plus two equals four' is contingent because people might have spoken a language in which 'two plus two equals four' meant that seven is even. Similarly, when we speak of a counterfactual situation, we speak of it in English, even if it is part of the description of that counterfactual situation that we were all speaking German in that counterfactual situation. We say, 'suppose we had all been speaking German' or 'suppose we

had been using English in a nonstandard way'. Then we are describing a possible world or counterfactual situation in which people, including ourselves, did speak in a certain way different from the way we speak. But still, in describing that world, we use *English* with *our* meanings and *our* references. It is in this sense that I speak of it as having the same reference in all possible worlds. I also don't mean to imply that the thing exists in all possible worlds, just that the name refers rigidly to that thing. If you say 'suppose Hitler had never been born' then 'Hitler' refers, here, still rigidly, to something that would not exist in the counterfactual situation described.

Given these remarks, ~~this means we must cross off Thesis (6) as incorrect.~~ The other theses have nothing to do with necessity and can survive. In particular Thesis (5) has nothing to do with necessity and it can survive. If I use the name 'Hesperus' to refer to a certain planetary body when seen in a certain celestial position in the evening, it will not therefore be a necessary truth that Hesperus is ever seen in the evening. That depends on various contingent facts about people being there to see and things like that. ~~So even if I should say to myself that I will use 'Hesperus' to name the heavenly body I see in the evening in yonder position of the sky, it will not be necessary that Hesperus was ever seen in the evening. But it may be *a priori* in that this is how I have determined the referent.~~ If I have determined that Hesperus is the thing that I saw in the evening over there, then I will know, just from making that determination of the referent, that if there is any Hesperus at all it's the thing I saw in the evening. This at least survives as far as the arguments we have given go so far.

How about a theory where Thesis (6) is eliminated? Theses (2), (3), and (4) turn out to have a large class of counterinstances. Even when Theses (2)–(4) are true, Thesis (5) is usually false; the truth of Theses (3) and (4) is an empirical 'accident', which the speaker hardly knows *a priori*. That is to say, other principles really determine the speaker's reference, and the fact that the referent coincides with that determined by (2)–(4) is an 'accident', which we were in no position to know *a priori*. Only in a rare class of cases, usually initial baptisms, are all of (2)–(5) true.

What picture of meaning do these Theses ((1)–(5)) give you? The picture is this. I want to name an object. I think of some way of describing it uniquely and then I go through, so to speak, a sort of mental

ceremony: By 'Cicero' I shall mean the man who denounced Cataline; and that's what the reference of 'Cicero' will be. I will use 'Cicero' to designate rigidly the man who (in fact) denounced Cataline, so I can speak of possible worlds in which he did not. But still my intentions are given by first, giving some condition which uniquely determines an object, then using a certain word as a name for the object determined by these conditions. Now there may be some cases in which we actually do this. Maybe, if you want to stretch and call it description, when you say: I shall call that heavenly body over there 'Hesperus'.³³ That is really a case where the theses not only are true but really even give a correct picture of how the reference is determined. Another case, if you want to call this a name, might be when the police in London use the name 'Jack' or 'Jack the Ripper' to refer to the man, whoever he is, who committed all these murders, or most of them. Then they are giving the reference of the name by a description.³⁴ But in many or most cases, I think the theses are false.³⁵ So let's look at them.

Thesis (1), as I say, is a definition. Thesis (2) says that one of the properties believed by *A* of the object or some conjointly are believed to pick out some individual uniquely. A sort of example people have in mind is just what I said: I shall use the term 'Cicero' to denote the man who denounced Cataline (or first denounced him in public, to make it unique). This picks out an object uniquely in this particular reference. Even some writers such as Ziff in *Semantic Analysis*, who don't believe that names have meaning in our sense, think that this is a good picture of the way reference can be determined.

Let's see if Thesis (2) is true. It seems, in some *a priori* way, that it's got to be true, because if you don't think that the properties you have in mind pick out any one uniquely – let's say they're all satisfied by two people – then how can you say which one of them you're talking about? There seem to be no grounds for saying you're talking about the one rather than about the other. Usually the properties in question are supposed to be some famous deeds of the person in question. For example, Cicero was the man who denounced Cataline. The average person, according to this, when he refers to Cicero, is saying something like 'the man who denounced Cataline' and thus has picked out a certain man uniquely. It is a tribute to the education of philosophers that they have held this thesis for such a long time. In fact, most people, when they

think of Cicero, just think of a famous Roman orator, without any pretension to think either that there was only one famous Roman orator or that one must know something else about Cicero to have a referent for the name. Consider Richard Feynman, to whom many of us are able to refer. He is a leading contemporary theoretical physicist. Everyone *here* (I'm sure!), can state the contents of one of Feynman's theories so as to differentiate him from Gell-Mann. However, the man in the street, not possessing these abilities, may still use the name 'Feynman'. When asked he will say: well he's a physicist or something. He may not think that this picks out anyone uniquely. I still think he uses the name 'Feynman' as a name for Feynman.

But let's look at some of the cases where we do have a description to pick out someone uniquely. Let's say, for example, that we know that Cicero was the man who first denounced Cataline. Well, that's good. That really picks someone out uniquely. However, there is a problem, because this description contains another name, namely 'Cataline'. We must be sure that we satisfy the conditions in such a way as to avoid the non-circularity condition here. In particular, we must not say that Cataline was the man denounced by Cicero. If that is so, we will really not be picking out anything uniquely, we will simply be picking out a pair of objects A and B , such that A denounced B or B denounced A . We do not think that this was the only pair where such denunciations ever occurred; so we had better add some other conditions in order to satisfy the uniqueness condition.

If we say Einstein was the man who discovered the theory of relativity, that certainly picks out someone uniquely. One can be sure, as I said, that everyone *here* can make a compact and independent statement of this theory and so pick out Einstein uniquely; but many people actually don't know enough about this stuff, so when asked what the theory of relativity is, they will say: 'Einstein's theory', and thus be led into the most straightforward sort of vicious circle.

So Thesis (2), in a straightforward way, fails to be satisfied when we say Feynman is a famous physicist without attributing anything else to Feynman. In another way it may not be satisfied in the proper way even when it is satisfied: If we say Einstein was 'the man who discovered relativity theory', that does pick someone out uniquely; but it may not pick him out in such a way as to satisfy the noncircularity condition,

because the theory of relativity may in turn be picked out as 'Einstein's theory'. So Thesis (2) seems to be false.

By changing the conditions ϕ from those usually associated with names by philosophers, one could try to improve the theory. There have been various ways I've heard; maybe I'll discuss these later on. Usually they think of famous achievements of the man named. Certainly in the case of famous achievements, the theory doesn't work. Some student of mine once said, 'Well Einstein discovered the theory of relativity'; and he determined the reference of 'the theory of relativity' independently by referring to an encyclopedia which would give the details of the theory. (This is what is called a transcendental deduction of the existence of encyclopedias.) But it seems to me that, even if someone has heard of encyclopedias, it really is not essential for his reference that he should know whether this theory is given in detail in any encyclopedia. The reference might work even if there had been no encyclopedias at all.

Let's go on to Thesis (3): If most of the ϕ 's, suitably weighted, are satisfied by a unique object y , then y is the referent of the name for the speaker. (Now, since we have already established that Thesis (2) is wrong, why should any of the rest work? The whole theory depended on always being able to specify unique conditions which are satisfied. But still we can look at the other theses. The picture associated with the theory is that only by giving some unique properties can you know who someone is and thus know what the reference of your name is. Well, I won't go into the question of knowing who someone is. It's really very puzzling. I think you *do* know who Cicero is if you just can answer that he's a famous Roman orator. Strangely enough, if you know that Einstein discovered the theory of relativity and nothing about that theory, you can both know who Einstein is, namely the discoverer of the theory of relativity, and who discovered the theory of relativity, namely Einstein, on the basis of this knowledge. This seems to be a blatant violation of some sort of noncircularity condition; but it is the way we talk. It therefore would seem that a picture which suggests this condition must be the wrong picture.)

Suppose most of the ϕ 's are in fact satisfied by a unique object. Is that object necessarily the referent of ' X ' for A ? Let's suppose someone says that Gödel is the man who proved the incompleteness of arithmetic, and this man is suitably well educated and is even able to give an in-

dependent account of the incompleteness theorem. He doesn't just say, 'Well, that's Gödel's theorem', or whatever. He actually states a certain theorem which he attributes to Gödel as the discoverer. Is it the case, then, that if most of the ϕ 's are satisfied by a unique object y then y is the referent of the name 'X' for A ? Let's take a simple case. In the case of Gödel that's practically the only thing many people have heard about him – that he discovered the incompleteness of arithmetic. Does it follow that whoever discovered the incompleteness of arithmetic is the referent of 'Gödel'?

Imagine the following blatantly fictional situation. (I hope Professor Gödel is not present.) Suppose that Gödel was not in fact the author of this theorem. A man named 'Schmidt', whose body was found in Vienna under mysterious circumstances many years ago, actually did the work in question. His friend Gödel somehow got hold of the manuscript and it was thereafter attributed to Gödel. On the view in question, then, when our ordinary man uses the name 'Gödel', he really means to refer to Schmidt, because Schmidt is the unique person satisfying the description, 'the man who discovered the incompleteness of arithmetic'. Of course you might try changing it to 'the man who *published* the discovery of the incompleteness of arithmetic'. By changing the story a little further one can make even this formulation false. Anyway most people might not even know whether the thing was published or got around by word of mouth. Let's stick to 'the man who discovered the incompleteness of arithmetic'. So, since the man who discovered the incompleteness of arithmetic is in fact Schmidt, we, when we talk about 'Gödel', are in fact always referring to Schmidt. But it seems to me that we are not. We simply are not. One reply, which I will discuss later, might be: You should say instead, 'the man to whom the incompleteness of arithmetic is commonly attributed', or something like that. Let's see what we can do with that later.

But it may seem to many of you that this is a very odd example, or that such a situation occurs rarely. This also is a tribute to the education of philosophers. Very often we use a name on the basis of considerable misinformation. The case of mathematics used in the fictive example is a good case in point. What do we know about Peano? What many people in this room may 'know' about Peano is that he was the discoverer of certain axioms which characterize the sequence of natural numbers, the

so-called 'Peano axioms'. Probably some people can even state them. I have been told that these axioms are not actually due to Peano but to Dedekind. Peano was of course not a dishonest man. He includes them in his book with an accompanying credit in his footnotes. Somehow the footnote has been ignored. So on the theory in question the term 'Peano', as we use it, really refers to – now that you've heard it you see that you were really all the time talking about – Dedekind. But you were not. Such illustrations could be multiplied indefinitely.

Even worse misconceptions, of course, occur to the layman. In a previous example I supposed people to identify Einstein by reference to his work on relativity. Actually, I often used to hear that Einstein's most famous achievement was the invention of the atomic bomb. So when we refer to Einstein we refer to the inventor of the atomic bomb. But this is not so. Columbus was the first man to realize that the earth was round. He was also the first European to land in the western hemisphere. Probably none of these things are true, and therefore, when people use the term 'Columbus' they really refer to some Greek if they use the roundness of the earth, or to some Norseman, perhaps, if they use the 'discovery of America'. But they don't. So it does not seem that if most of the ϕ 's are satisfied by a unique object y , then y is the referent of the name. This seems simply to be false.³⁶

Thesis (4): If the vote yields no unique object the name does not refer. Really this case has been covered before – has been covered in my previous examples. First, the vote may not yield a *unique* object, as in the case of Cicero or Feynman. Secondly, suppose it yields *no* object, that nothing satisfies most, or even any, substantial number, of the ϕ 's. Does that mean the name doesn't refer? No: in the same way that you may have false beliefs about a person which may actually be true of someone else, so you may have false beliefs which are true of absolutely no one. And these may constitute the totality of your beliefs. Suppose, to vary the example about Gödel, no one had discovered the incompleteness of arithmetic – perhaps the proof simply materialized by a random scattering of atoms on a piece of paper – the man Gödel being lucky enough to have been present when this improbable event occurred. Further, suppose arithmetic is in fact complete. One wouldn't really expect a random scattering of atoms to produce a correct proof. A subtle error, unknown through the decades, has still been unnoticed – or perhaps

not actually unnoticed, but ^{only to} the friends of Gödel So even if the conditions are not satisfied by a unique object the name may still refer. I gave you the case of Jonah last week. Biblical scholars, as I said, think that Jonah really existed. It isn't because they think that someone ever was swallowed by a big fish or even went to Nineveh to preach. These conditions may be true of no one whatsoever and yet the name 'Jonah' really has a referent. In the case above of Einstein's invention of the bomb, possibly no one really deserves to be called the 'inventor' of the device.

Thesis 5 says that the statement 'If X exists, then X has most of the ϕ 's', is *a priori* true for A . Notice that even in a case where (3) and (4) *happen* to be true, a typical speaker hardly knows *a priori* that they are, as required by the theory. I *think* that my belief about Gödel *is* in fact correct and that the 'Schmidt' story is just a fantasy. But the belief hardly constitutes *a priori* knowledge.

What's going on here? Can we rescue the theory?³⁷ First, one may try and vary these descriptions – not think of the famous achievements of a man but, let's say, of something else, and try and use that as our description. Maybe by enough futzing around someone might eventually get something out of this³⁸; however, most of the attempts that one tries are open to counterexamples or other objections. Let me give an example of this. In the case of Gödel one may say, 'Well, 'Gödel' doesn't mean 'the man who proved the incompleteness of arithmetic''. Look, all we really know is that most people *think* that Gödel proved the incompleteness of arithmetic, that Gödel is the man to whom the incompleteness of arithmetic is commonly attributed. So when I determine the referent of the name 'Gödel', I don't say to myself, 'by 'Gödel' I shall mean 'the man who proved the incompleteness of arithmetic, whoever he is''. That might turn out to be Schmidt or Post. But instead I shall mean 'the man who most people *think* proved the incompleteness of arithmetic'.

Is this right? First, it seems to me that it's open to counterexamples of the same type as I gave before, though the counterexamples may be more *récherché*. Suppose, in the case of Peano mentioned previously, unbeknownst to the speaker most people (at least by now) thoroughly realize that the number-theoretic axioms should not be attributed to him. Most people don't credit them to Peano but now correctly ascribe them to Dedekind. So then even the man to whom this thing is commonly

attributed will still be Dedekind and not Peano. Still, the speaker, having picked up the old outmoded belief, may still be referring to Peano, and hold a false belief about Peano, not a true belief about Dedekind.

But second, and perhaps more significantly, such a criterion violates the noncircularity condition. How is this? It is true that most of us think that Gödel proved the incompleteness of arithmetic. Why is this so? We certainly say, and sincerely, 'Gödel proved the incompleteness of arithmetic'. Does it follow from that that we believe that Gödel proved the incompleteness of arithmetic – that we attribute the incompleteness of arithmetic to this man? No. Not just from that. We have to be *referring* to Gödel when we say 'Gödel proved the incompleteness of arithmetic'. If, in fact, we were always referring to Schmidt, then we would be attributing the incompleteness of arithmetic to Schmidt and not to Gödel – if we used the sound 'Gödel' as the name of the man whom I am calling 'Schmidt'.

But we do in fact refer to Gödel. How do we do this? Well, not by saying to ourselves, 'By 'Gödel' I shall mean the man to whom the incompleteness of arithmetic is commonly attributed'. If we did that we would run into a circle. Here we are all in this room. Actually in this institution³⁹ some people have met the man, but in many institutions this is not so. All of us in the community are trying to determine the reference by saying 'Gödel is to be the man to whom the incompleteness of arithmetic is commonly attributed'. None of us will get started with any attribution unless there is some independent criterion for the reference of the name other than 'the man to whom the incompleteness of arithmetic is commonly attributed'. Otherwise all we will be saying is, 'We attribute this achievement to the man to whom we attribute it', without saying who that man is, without giving any independent criterion of the reference, and so the determination will be circular. This then is a violation of the condition I have marked 'C', and cannot be used in any theory of reference.

Of course you might try to avoid circularity by passing the buck. This is mentioned by Strawson, who says in his footnote on these matters that one man's reference may derive from another's.

The identifying description, though it must not include a reference to the speaker's own reference to the particular in question, may include a reference to another's reference to that particular. If a putatively identifying description is of this latter kind,

then, indeed, the question, whether it is a genuinely identifying description turns on the question, whether the reference it refers to is itself a genuinely identifying reference. So one reference may borrow its credentials, as a genuinely identifying reference, from another; and that from another. But this regress is not infinite.⁴⁰

I may then say, 'Look, by 'Gödel' I shall mean the man Joe thinks proved the incompleteness of arithmetic'. Joe may then pass the thing over to Harry. One has to be very careful that this doesn't come round in a circle. Is one really sure that this won't happen? If you could be sure yourself of knowing such a chain, and that everyone else in the chain is using the proper conditions and so is not getting out of it, then maybe you could get back to the man by referring to such a chain in that way, borrowing the references one by one. However, although in general such chains do exist for a living man, you won't know what the chain is. You won't be sure what descriptions the other man is using, so the thing won't go into a circle, or whether by appealing to Joe you won't get back to the right man at all. So you cannot use this as your identifying description with any confidence. You may not even remember from whom you heard of Gödel.

What is the true picture of what's going on? Maybe reference doesn't really take place at all! After all, we don't really know that any of the properties we use to the man are right. We don't know that they pick out a unique object. So what *does* make my use of 'Cicero' into a name of *him*? The picture which leads to the cluster-of descriptions theory is something like this: One is isolated in a room; the entire community of other speakers, everything else, could disappear; and one determines the reference for himself by saying - 'By 'Gödel' I shall mean the man, whoever he is, who proved the incompleteness of arithmetic'. Now you can do this if you want to. There's nothing really preventing it. You can just stick to that determination. If that's what you do, then if Schmidt discovered the incompleteness of arithmetic you *do* refer to him when you say 'Gödel did such and such'.

But that's not what most of us do. Someone, let's say, a baby, is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say Richard Feynman, in the market place or elsewhere, may be referring to Richard Feynman even

chain

though he can't remember from whom he first heard of Feynman or from whom he ever heard of Feynman. He knows that Feynman was a famous physicist. A certain passage of communication reaching ultimately to the man himself does reach the speaker. He then is referring to Feynman even though he can't identify him uniquely. He doesn't know what a Feynman diagram is, he doesn't know what the Feynman theory of pair production and annihilation is. Not only that: he'd have trouble distinguishing between Gell-Mann and Feynman. So he doesn't have to know these things, but, instead, a chain of communication going back to Feynman himself has been established, by virtue of his membership in a community which passed the name on from link to link, not by a ceremony that he makes in private in his study: 'By 'Feynman' I shall mean the man who did such and such and such and such'.

How does this view differ from Strawson's suggestion, mentioned before, that one identifying reference may borrow its credentials from another? Certainly Strawson had a good insight in the passage quoted; on the other hand, he certainly shows a difference at least in emphasis from the picture I advocate, since he confines the remark to a footnote. The main text advocates the cluster-of-descriptions theory. Just because Strawson makes his remark in the context of a description theory, his view therefore differs from mine in one important respect. Strawson apparently must require that the speaker must *know* from whom he got his reference, so that he can say: 'By 'Gödel' I mean the man *Jones* calls 'Gödel''. If he does not remember how he picked up the reference, he cannot give such a description. The present theory sets no such requirement. As I said, I may well not remember from whom I heard of Gödel, and I may think I remember from which people I heard the name, but wrongly.

These considerations show that the view advocated here can lead to consequences which actually *diverge* from those of Strawson's footnote. Suppose that the speaker has heard the name 'Cicero' from Smith and others, who use the name to refer to a famous Roman orator. He later thinks, however, that he picked up the name from Jones, who (unknown to the speaker) uses 'Cicero' as the name of a notorious German spy and has never heard of any orators of the ancient world. Then, according to Strawson's paradigm, the speaker must determine his reference by the resolution, 'I shall use 'Cicero' to refer to the man whom Jones calls

by that name', while on the present view, the referent will be the orator in spite of the speaker's false impression about where he picked up the name. The point is that Strawson, trying to fit the chain of communication view into the description theory, is forced to rely on what the speaker *thinks* was the source of his reference. If the speaker has forgotten his source, the device is unavailable to Strawson; if he misremembers it, Strawson's paradigm in his footnote can give the wrong results. On our view, it is not how the speaker thinks he got the reference, but the actual chain of communication, which is relevant.

I think I said the other time that philosophical theories are in danger of being false, and so I wasn't going to present an alternative theory. Have I just done so? Well, in a way; but my characterization has been far less specific than a real set of necessary and sufficient conditions for reference would be. Obviously the name is passed on from link to link. But of course not every sort of causal chain reaching from me to a certain man will do for me to make a reference. There may be a causal chain from our use of the term 'Santa Claus' to a certain historical saint, but still the children, when they use this, by this time probably do not refer to that saint. So other conditions must be satisfied in order to make this into a really rigorous theory of reference. I don't know that I'm going to do this because, first, I'm sort of too lazy at the moment; secondly, rather than giving a set of necessary and sufficient conditions which will work for a term like reference, I want to present just a *better picture* than the picture presented by the received views.

Haven't I been very unfair to the description theory? Here I have stated it very precisely – more precisely, perhaps, than it has been stated by any of its advocates. So then it's easy to refute. Maybe if I tried to state mine with sufficient precision in the form of six or seven or eight theses, it would also turn out that when you examine the theses one by one, they will all be false. That might even be so, but the difference is this. ~~What I think the examples I've given show is not simply that there's some technical error here or some mistake there, but that the whole picture given by this theory of how reference is determined seems to be wrong from the fundamentals. It seems to be wrong to think that we give ourselves some properties which somehow qualitatively uniquely pick out an object and determine our reference in that manner. What I am trying to present is a better picture – a picture which, if more details were~~

to be filled in, might be refined so as to give more exact conditions for reference to take place.

One might never reach a set of necessary and sufficient conditions. I don't know, I'm always sympathetic to Bishop Butler's 'Everything is what it is and not another thing' – in the nontrivial sense that philosophical analyses of some concept like reference, in completely different terms which make no mention of reference, are very apt to fail. Of course in any particular case when one is given an analysis one has to look at it and see whether it is true or false. One can't just cite this maxim to oneself and then turn the page. But more cautiously, I want to present a better picture without giving a set of necessary and sufficient conditions for reference. Such conditions would be very complicated, but what is true is that it's in virtue of our connection with other speakers in the community, going back to the referent himself, that we refer to a certain man.

There may be some cases where the description picture is true, where some man really gives a name by going into the privacy of his room and saying that the referent is to be the unique thing with certain identifying properties. 'Jack the Ripper' was a possible example which I gave. Another was 'Hesperus'. Yet another case which can be forced into this description is that of meeting someone and being told his name. Except for a belief in the description theory, in its importance in other cases, one probably wouldn't think that that was a case of giving oneself a description, i.e., 'the guy I'm just meeting now'. But one can put it in these terms if one wishes, and if one has never heard the name in any other way. Of course, if you're introduced to a man and told, 'That's Einstein', you've heard of him before, it may be wrong, and so on. But maybe in some cases such a paradigm works – especially for the man who first gives someone or something a name. Or he points to a star and says, 'That is to be Alpha Centauri'. So he can really make himself this ceremony: 'By 'Alpha Centauri' I shall mean the star right over there with such and such coordinates'. But in general this picture fails. In general our reference depends not just on what we think ourselves, but on other people in the community, the history of how the name reached one, and things like that. It is by following such a history that one gets to the reference.

More exact conditions are very complicated to give. They seem in a

STAMPED OFFICIAL

302

SAUL A. KRIPKE

way somehow different in the case of a famous man and one who isn't so famous.⁴¹ For example, a teacher tells his class that Newton was famous for being the first man to think there's a force pulling things to the earth; and I think that's what little kids think Newton's greatest achievement was. I won't say what the merits of such an achievement would be, but, anyway, we may suppose that just being told that this was the sole content of Newton's discovery gives the students a false belief *about Newton*, even though they have never heard of him before. If on the other hand the teacher uses the name, George Smith, who is in fact his next door neighbor, and says that George Smith first squared the circle, does it follow from this that the students have a false belief about the teacher's neighbor? The teacher doesn't tell them that Smith is his neighbor, nor does he believe Smith first squared the circle. He isn't particularly trying to get any belief *about the neighbor* into the students' heads. He tries to inculcate the belief that there was a man who squared the circle, but not a belief about any particular man – he just pulls out the first name that occurs to him – as it happens, he uses his neighbor's name. It doesn't seem clear in that case that the students have a false belief about the neighbor, even though there is a causal chain going back to the neighbor. I am not sure about this. At any rate more refinements need to be added to make this even begin to be a set of necessary and sufficient conditions. In that sense it's not a theory, but is supposed to give a better picture of what is actually going on.

~~A rough statement of a theory might be the following: An initial baptism takes place. Here the object may be named by ostension, or the reference of the same may be fixed by a description.⁴² When the name is 'passed from link to link', the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it. If I hear the name 'Napoleon' and decide it would be a nice name for my pet aardvark, I do not satisfy this condition.⁴³ (Perhaps it is some such failure to keep the reference fixed which accounts for the divergence of present uses of 'Santa Claus' from the alleged original use.)~~

Notice that the preceding outline hardly *eliminates* the notion of reference; on the contrary, it takes the notion of intending to use the same reference as a given. This is also an appeal to an initial baptism which is explained in terms either of fixing a reference by a description, or ostension (if ostension is not to be subsumed under the other category).⁴⁴

Necessity & identity

(Perhaps there are other possibilities for initial baptisms.) Further, the George Smith case casts some doubt as to the sufficiency of the conditions. Even if the teacher does refer to his neighbor, is it clear that he has passed on his reference to the pupils? Why shouldn't their belief be about any other man named 'George Smith'? If he says that Newton was hit by an apple, somehow his task of transmitting a reference is easier, since he has communicated a common misconception about Newton.

To repeat, I may not have presented a theory, but I do think that I have presented a better picture than that given by description theorists.

I think the next topic I shall want to talk about is that of statements of identity. Are these necessary or contingent? The matter has been in some dispute in recent philosophy. First, everyone agrees that descriptions can be used to make contingent identity statements. If it is true that the man who invented bifocals was the first postmaster general of the United States – that these were one and the same – it's contingently true. That is, it might have been the case that one man invented bifocals and another was the first postmaster general of the United States. So certainly when you make identity statements using descriptions – when you say 'the x such that ϕx and the x such that ψx are one and the same', that can be a contingent fact. But philosophers have been interested also in the question of identity statements between names. When we say 'Hesperus is Phosphorus' or 'Cicero is Tully', is what we are saying necessary or contingent? Further, they've been interested in another type of identity statement, which comes from scientific theory. We identify, for example, light with electromagnetic radiation between certain limits of wavelengths, or with a stream of photons. We identify heat with the motion of molecules; sound with a certain sort of wave disturbance in the air; and so on. Concerning such statements the following thesis is commonly held. First, that these are obviously contingent identities: we've found out that light is a stream of photons, but of course it might not have been a stream of photons. Heat is in fact the motion of molecules; we found that out, but heat might not have been the motion of molecules. Secondly, many philosophers feel damned lucky that these examples are around. Now, why? These philosophers, whose views are expounded in a vast literature, hold to a thesis called the identity thesis with respect to some

Heat = motion of U - truth

SAUL A. KRIPKE

psychological concepts. They think, say, that pain is just a certain material state of the brain or of the body, or what have you – say the stimulation of C-fibres. (It doesn't matter what.) Some people have then objected, "Well, look, there's perhaps a *correlation* between pain and these states of the body; but this must just be a contingent correlation between two different things, because it was an empirical discovery that this correlation ever held. Therefore, by 'pain' we must mean something different from this state of the body or brain; and, therefore, they must be two different things."

Then it's said, "Ah, but you see, this is wrong! Everyone knows that there can be contingent identities." First, as in the bifocals and post-master general case, which I have mentioned before. Second, in the case, believed closer to the present paradigm, of theoretical identifications, such as light and a stream of photons, or water and a certain compound of hydrogen and oxygen. These are all contingent identities. They might have been false. It's no surprise, therefore, that it can be true as a matter of contingent fact and not of any necessity that feeling pain, or seeing red, is just a certain state of the human body. Such psychophysical identifications can be contingent facts just as the other identities are contingent facts. And of course there are widespread motivations – ideological, or just not wanting to have the nomological dangler of these mysterious connections not accounted for by laws of physics, one to one correlation between two different things – a material state and something of an entirely different kind – which lead people to want to believe this thesis.

I guess the main thing I'll talk about first is identity statements between names. But I hold the following about the general case. First, that characteristic theoretical identifications like 'Heat is the motion of molecules', are not contingent truths but necessary truths, and here of course I don't mean just physically necessary, but necessary in the highest degree – whatever that means. (Physical necessity, might turn out to be necessity in the highest degree. But that's a question which I don't wish to prejudge. At least for this sort of example, it might be that when something's physically necessary, it always is necessary *tout court*.) Second, that the way in which these have turned out to be necessary truths does not seem to me to be a way in which the mind-brain identities could turn out to be either necessary or contingently true. So this analogy has to go.

Identities & names are necessary

It's hard to see what to put in its place. It's hard to see therefore how to avoid concluding that the two are actually different.

Let me go back to the more mundane case about proper names. This is already mysterious enough. There's a dispute about this between Quine and Ruth Barcan Marcus.⁴⁵ Ruth Barcan Marcus says that identities between names are necessary. If someone thinks that Cicero is Tully, and really uses 'Cicero' and 'Tully' as names, he is thereby committed to holding that his belief is a necessary truth. She uses the term 'mere tag'. Quine replies as follows, "We may tag the planet Venus, some fine evening, with the proper name 'Hesperus'. We may tag the same planet again, some day before sunrise, with the proper name 'Phosphorus'. When we discover that we have tagged the same planet twice our discovery is empirical. And not because the proper names were descriptions."⁴⁶ First, Quine is right when we discovered that we tagged the same planet twice, our discovery was empirical. Another example I think Quine gives in another book is that the same mountain seen from Nepal and from Tibet, or something like that, is from one angle called Mt. Everest (you've heard of that); from another it's supposed to be called Gaurisanker. It can actually be an empirical discovery that Gaurisanker is Everest. (Quine says that the example is actually false. He got the example from Erwin Schrodinger. You wouldn't think the inventor of wave mechanics got things that wrong. I don't know where the mistake is supposed to come from. One could certainly imagine this situation as having been the case; and it's another good illustration of the sort of thing that Quine has in mind.)

What about it? I wanted to find a good quote on the other side from Mrs. Marcus in this book but I am having trouble locating one. Being present at that discussion, I remember⁴⁷ she advocated the view that if you really have names, a good dictionary should be able to tell you whether they have the same reference. So someone should be able, by looking in the dictionary, to say that Hesperus and Phosphorus are the same. Now this does not seem to be true. It does seem, to many people, to be a consequence of the view that identities between names are necessary. Therefore the view that identity statements between names are necessary has usually been rejected. Russell's conclusion was somewhat different. He did think there should never be any empirical question whether two names have the same reference. This isn't satisfied for

because of the way
 reason to think so

ordinary names, but it is satisfied when you're naming your own sense datum, or something like that. You say, "Here, this, and that (designating the same sense datum by both demonstratives)." So you can tell that you're naming the same thing twice; the conditions are satisfied. Since this won't apply to ordinary cases of naming, ordinary 'names' cannot be genuine names.

What should we think about this? First, it's true that someone can use the name 'Cicero' to refer to Cicero and the name 'Tully' to refer to Cicero also, and not know that Cicero is Tully. So it seems that we do not necessarily know *a priori* that an identity statement between names is true. It doesn't follow from this that the statement so expressed is a contingent one if true. This is what I've emphasized in my first lecture. There are very strong feeling which lead one to think that, if you can't know something by *a priori* ratiocination, then it's got to be contingent: it might have turned out otherwise; but nevertheless I think this feeling is wrong.

Let's suppose we refer to the same heavenly body twice, as Hesperus and Phosphorus. We say Hesperus is that star over there in the evening; Phosphorus is that star over there in the morning. Actually Hesperus is Phosphorus. Are there really circumstances under which Hesperus wouldn't have been Phosphorus? Supposing that Hesperus is Phosphorus, let's try to describe a possible situation in which it would not have been. Well it's easy. Someone goes by and he calls two *different* stars Hesperus and Phosphorus. It may be even under the same conditions as prevailed when we introduced the names 'Hesperus' and 'Phosphorus'. But are those circumstances in which Hesperus is not Phosphorus or would not have been Phosphorus? It seems to me that they are not.

Now, of course I'm committed to saying that they're not, by saying that such terms as 'Hesperus' and 'Phosphorus', when used as names, are rigid designators. They refer in every possible world to the planet Venus. Therefore, in that possible world too, the planet Venus is the planet Venus and it doesn't matter what any other person has said in this other possible world. How should we describe this situation? He can't have pointed to Venus twice, and in the one case called it Hesperus and in the other Phosphorus, as we did. If he did so, then 'Hesperus is Phosphorus' would have been true in that situation too. He pointed maybe neither time to the planet Venus - at least one time he didn't point to the planet Venus, let's say when he pointed to Phosphorus. Then in that case we

can certainly say that the name 'Phosphorus' might not have referred to Phosphorus. We can even say that in the very position when viewed in the morning that we found Phosphorus, it might have been the case that Phosphorus was not there – that something else was there, and that even, under certain circumstances it would have been *called* 'Phosphorus'. But that still is not a case in which Phosphorus was not Hesperus. There might be a possible world in which, a possible counterfactual situation in which, 'Hesperus' and 'Phosphorus' weren't names of the things they in fact are names of. Someone, if he did determine their reference by identifying descriptions, might even have used the very identifying descriptions we used. But still that's not a case in which Hesperus wasn't Phosphorus. For there couldn't have been such a case, given that Hesperus is Phosphorus.

Now this seems very strange, because in advance, we are inclined to say, the answer to the question whether Hesperus is Phosphorus might have turned out either way. So aren't there really two possible worlds – one in which Hesperus was Phosphorus, the other in which Hesperus wasn't Phosphorus – in advance of our discovering that these were the same? First, there's one sense in which things might turn out either way, in which it's clear that that doesn't imply that the way it finally turns out isn't necessary. For example, the four color theorem might turn out to be true and might turn out to be false. It might turn out either way. It still doesn't mean that the way it turns out is not necessary. Obviously, the 'might' here is purely 'epistemic' – it merely expresses our present state of ignorance, or uncertainty.

But it seems that in the Hesperus-Phosphorus case, something even stronger is true. The evidence I have before I know that Hesperus is Phosphorus is that I see a certain star or a certain heavenly body in the evening and call it Hesperus, and in the morning and call it Phosphorus. I know these things. There certainly is a possible world in which a man should have seen a certain star at a certain position in the evening and called it 'Hesperus' and a certain star in the morning and called it 'Phosphorus'; and should have concluded – should have found out by empirical investigation – that he names two different stars, or two different heavenly bodies. At least one of these stars or heavenly bodies was not Phosphorus, otherwise it couldn't have come out that way. But that's true. And so it's true that given the evidence that someone has antecedent

to his empirical investigation, he can be placed in a sense in exactly the same situation, that is a qualitatively identical epistemic situation, and call two heavenly bodies 'Hesperus' and 'Phosphorus', without their being identical. So in that sense we can say that it might have turned out either way. Not that it might have turned out either way as to Hesperus's being Phosphorus. Though for all we knew in advance, Hesperus wasn't Phosphorus, that couldn't have turned out any other way, in a sense. But being put in a situation where we have exactly the same evidence, qualitatively speaking, it could have turned out that Hesperus was not Phosphorus; that is, in a counterfactual world in which 'Hesperus' and 'Phosphorus' were not used in the way that we use them, as names of this planet, but as names of some other objects, one could have had qualitatively identical evidence and concluded that 'Hesperus' and 'Phosphorus' named two different objects.⁴⁸ But we, using the names as we do right now, can say in advance, that if Hesperus and Phosphorus are one and the same then in no other possible world can they be different. We use 'Hesperus' as the name of a certain body and 'Phosphorus' as the name of a certain body. We use it as the name of those bodies in all possible worlds. If, in fact, they are the *same* body, then in any other possible world we have to use it as a name of that object. And so in any other possible world it will be true that Hesperus is Phosphorus. So two things are true: first, that we do not know *a priori* that Hesperus is Phosphorus, and are in no position to find out the answer except empirically. Second, this is so because we would have evidence qualitatively undistinguishable from the evidence we have and determine the reference of two names by the positions of two planets in the sky, without the planets being the same.

Of course, it won't be true in every other possible world that the star seen over there in the evening is the star seen over there in the morning, because there might have been possible worlds in which Phosphorus was not visible in the morning. But that shouldn't be identified with the statement that Hesperus is Phosphorus. It could only be so identified if you thought that it was a necessary truth that Hesperus is visible over there in the evening or that Phosphorus is visible over there in the morning. But neither of those are necessary truths even if that's the way we pick out the planet. These are the contingent marks by which we identify a certain planet and give it a name.