Selections from

NAMING AND NECESSITY

Saul A. Kripke

Harvard University Press
Cambridge, Massachusetts

FROM LECTURE 1

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 is 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 every possible world it designates the same object, a *nonrigid* or *accidental designator* if that is 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

[15] Misleadingly, because the phrase suggests that there is a special problem of 'transworld identification', that we cannot trivially stipulate whom or what we are talking about when we imagine another possible world. The term 'possible world' may also mislead; perhaps it suggests the 'foreign country' picture. I have sometimes used 'counterfactual situation' in the text; Michael Slote has suggested that 'possible state (or history) of the world' might be less misleading than 'possible world'. It is better still, to avoid confusion, not to say, 'In some possible world, Humphrey would have won' but rather, simply, 'Humphrey might have won'. The apparatus of possible words has (I hope) been very useful as far as the set-theoretic model-theory of quantified modal logic is concerned, but has encouraged philosophical pseudo-problems and misleading pictures.

[16] Of course I don't imply that language contains a name for every object. Demonstratives can be used as rigid designators, and free variables can be used as rigid designators of unspecified objects. Of course when we specify a counterfactual situation, we do not describe the whole possible world, but only the portion which interests us.

need not be given purely qualitatively, as if we were looking at them through a telescope. And we will see shortly that the properties an object has in every counterfactual world have nothing to do with properties used to identify it in the actual world.[17]

[17] See Lecture I, p. 53 (on Nixon), and Lecture II, pp. 74–7.

FROM LECTURE 2

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

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 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-fibers. (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

SELECTIONS from KRIPKE, NAMING & NECESSITY, PAGE 3

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 Postmaster 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 mysterious connections not accounted for by the laws of physics, one to one correlations between two different kinds of thing, material states, and things 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. 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.

FROM LECTURE 3

According to the view I advocate, then, terms for natural kinds are much closer to proper names than is ordinarily supposed. The old term 'common name' is thus quite appropriate for predicates marking out species or natural kinds, such as 'cow' or 'tiger'. My considerations apply also, however, to certain mass terms for natural kinds, such as 'gold', 'water', and the like. It is interesting to compare my views to those of Mill. Mill counts both predicates like 'cow', definite descriptions, and proper names as names. He says of 'singular' names that they are connotative if they are definite descriptions but non-connotative if they are proper names. On the other hand, Mill says that *all* 'general' names are connotative; such a predicate as 'human being' is defined as the conjunction of certain properties which give necessary and sufficient conditions for humanity—rationality, animality, and certain physical features.[65] The modern logical tradition, as represented by Frege and Russell, seems to hold that Mill was wrong about singular names, but right about general names. More recent philosophy has followed suit, except that, in the case of both proper names and natural kind terms, it often replaces the notion of defining properties by that of a cluster of properties, only some of which need to be satisfied in each particular case. My own view, on the other hand, regards Mill as more-or-less right about 'singular' names, but wrong about 'general' names. *Perhaps* some 'general' names ('foolish', 'fat', 'yellow') express

[65] Mill, op. cit.

SELECTIONS from KRIPKE, NAMING & NECESSITY, PAGE 4

properties.[66] In a significant sense, such general names as 'cow' and 'tiger' do not, unless *being a cow* counts trivially as a property. Certainly 'cow' and 'tiger' are *not* short for the conjunction of properties a dictionary would take to define them, as Mill thought. Whether science can discover empirically that certain properties are *necessary* of cows, or of tigers, is another question, which I answer affirmatively.

Let's consider how this applies to the types of identity statements expressing scientific discoveries that I talked about before—say, that water is H_2O. It certainly represents a discovery that water is H_2O. We identified water originally by its characteristic feel, appearance and perhaps taste, (though the taste may usually be due to the impurities). If there were a substance, even actually, which had a completely different atomic structure from that of water, but resembled water in these respects, would we say that some water wasn't H_2O? I think not. We would say instead that just as there is a fool's gold there could be a fool's water; a substance which, though having the properties by which we originally identified water, would not in fact be water. And this, I think, applies not only to the actual world but even when we talk about counterfactual situations. If there had been a substance, which was a fool's water, it would then be fool's water and not water. On the other hand if this substance can take another form—such as the polywater allegedly discovered in the Soviet Union, with very different identifying marks from that of what we now call water—it is a form of water because it is the same substance, even though it doesn't have the appearances by which we originally identified water.

Let's consider the statement 'Light is a stream of photons' or 'Heat is the motion of molecules'. By referring to light, of course, I mean something which we have some of in this room. When I refer to heat, I refer not to an internal sensation that someone may have, but to an external phenomenon which we perceive through the sense of feeling; it produces a characteristic sensation which we call the sensation of heat. Heat *is* the motion of molecules. We have also discovered that increasing heat corresponds to increasing motion of molecules, or, strictly speaking, increasing average kinetic energy of molecules. So temperature is identified with mean molecular kinetic energy. However I won't talk about temperature because there is the question of how the actual scale is to be set. It might just be set in terms of the mean molecular kinetic energy.[67] But what represents an interesting phenomenological discovery is that when it's hotter the molecules are moving faster. We have also discovered about light that light is a stream of photons; alternatively it is a form of electromagnetic radiation. Originally we identified light by the characteristic internal visual impressions it can produce in us, that make us able to see. Heat, on the other hand, we originally identified by the characteristic effect on one aspect of our nerve endings or our sense of touch.

Imagine a situation in which human beings were blind or their eyes didn't work. They were unaffected by light. Would that have been a situation in which light did not exist? It seems

[66] I am not going to give any criterion for what I mean by a 'pure property', or Fregean intension. It is hard to find unquestionable examples of what is meant. Yellowness certainly expresses a manifest physical property of an object and, relative to the discussion of gold above, can be regarded as a property in the required sense. Actually, however, it is not without a certain referential element of its own, for on the present view yellowness is picked out and rigidly designated as that external physical property of the object which we sense by means of the *visual impression of yellowness*. It does in this respect resemble the natural kind terms. The phenomenological quality of the sensation itself, on the other hand, can be regarded as a *quale* in some pure sense. Perhaps I am rather vague about these questions, but further precision seems unnecessary here.

[67] Of course, there is the question of the relation of the statistical mechanical notion of temperature to, for example, the thermodynamic notion. I wish to leave such questions aside in this discussion.

to me that it would not. It would have been a situation in which our eyes were not sensitive to light. Some creatures may have eyes not sensitive to light. Among such creatures are unfortunately some people, of course; they are called 'blind'. Even if all people had had awful vestigial growths and just couldn't see a thing, the light might have been around; but it would not have been able to affect people's eyes in the proper way. So it seems to me that such a situation would be a situation in which there was light, but people could not see it. So, though we may identify light by the characteristic visual impressions it produces in us, this seems to be a good example of fixing a reference. We fix what light is by the fact that it is whatever, out in the world, affects our eyes in a certain way. But now, talking about counterfactual situations in which let's say, people were blind, we would not then say that since, in such situations, nothing could affect their eyes, light would not exist; rather we would say that that would be a situation in which light—the thing we have identified as that which in fact enables us to see—existed but did not manage to help us see due to some defect in us.

Perhaps we can imagine that, by some miracle, sound waves somehow enabled some creature to see. I mean, they gave him visual impressions just as we have, maybe exactly the same color sense. We can also imagine the same creature to be completely *insensitive* to light (photons). Who knows what subtle undreamt of possibilities there may be? Would we say that in such a possible world, it was sound which was light, that these wave motions in the air were light? It seems to me that, given our concept of light, we should describe the situation differently. It would be a situation in which certain creatures, maybe even those who were called 'people' and inhabited this planet, were sensitive not to light but to sound waves, sensitive to them in exactly the same way that we are sensitive to light. If this is so, once we have found out what light is, when we talk about other possible worlds we are talking about *this* phenomenon in the world, and not using 'light' as a phrase *synonymous* with 'whatever gives us the visual impression—whatever helps us to see'; for there might have been light and it not helped us to see; and even something else might have helped us to see. The way we identified light *fixed a reference*.

And similarly for other such phrases, such as 'heat'. Here heat is something which we have identified (and fixed the reference of its name) by its giving a certain sensation, which we call 'the sensation of heat'. We don't have a special name for this sensation other than as a sensation of heat. It's interesting that the language is this way. Whereas you might suppose it, from what I am saying, to have been the other way. At any rate, we identify heat and are able to sense it by the fact that it produces in us a sensation of heat. It might here be so important to the concept that its reference is fixed in this way, that if someone else detects heat by some sort of instrument, but is unable to feel it, we might want to say, if we like, that the concept of heat is not the same even though the referent is the same.

Nevertheless, the term 'heat' doesn't *mean* 'whatever gives people these sensations'. For first, people might not have been sensitive to heat, and yet the heat still have existed in the external world. Secondly, let us suppose that somehow light rays, because of some difference in their nerve endings, *did* give them these sensations. It would not then be heat but light which gave people the sensation which we call the sensation of heat.

Can we then imagine a possible world in which heat was not molecular motion? We can imagine, of course, having discovered that it was not. It seems to me that any case which someone will think of, which he thinks at first is a case in which heat—contrary to what is actually the case—would have been something other than molecular motion, would actually be a case in which some creatures with different nerve endings from

ours inhabit this planet (maybe even we, if it's a contingent fact about us that we have this particular neural structure), and in which these creatures were sensitive to that something else, say light, in such a way that they felt the same thing that we feel when we feel heat. But this is not a situation in which, say, light would have been heat, or even in which a stream of photons would have been heat, but a situation in which a stream of photons would have produced the characteristic sensations which *we* call 'sensations of heat'.

Similarly for many other such identifications, say, that lightning is electricity. Flashes of lightning are flashes of electricity. Lightning is an electrical discharge. We can imagine, of course, I suppose, other ways in which the sky might be illuminated at night with the same sort of flash without any electrical discharge being present. Here too, I am inclined to say, when we imagine this, we imagine something with all the visual appearances of lightning but which is not, in fact, lightning. One could be told: this appeared to be lightning but it was not. I suppose this might even happen now. Someone might, by a clever sort of apparatus, produce some phenomenon in the sky which would fool people into thinking that there was lightning even though in fact no lightning was present. And you wouldn't say that that phenomenon, because it looks like lightning, was in fact lightning. It was a different phenomenon from lightning, which is the phenomenon of an electrical discharge; and this is not lightning but just something that deceives us into thinking that there is lightning.

Usually, when a proper name is passed from link to link, the way the reference of the name is fixed is of little importance to us. It matters not at all that different speakers may fix the reference of the name in different ways, provided that they give it the same referent. The situation is probably not very different for species names, though the temptation to think that the metallurgist has a different concept of gold from the man who has never seen any may be somewhat greater. The interesting fact is that the way the reference is fixed seems overwhelmingly important to us in the case of sensed phenomena: a blind man who uses the term 'light', even though he uses it as a rigid designator for the very same phenomenon as we, seems to us to have lost a great deal, perhaps enough for us to declare that he has a different concept. ('Concept' here is used non-technically!) The fact that we identify light in a certain way seems to us to be *crucial*, even though it is not necessary; the intimate connection may create an *illusion* of necessity. I think that this observation, together with the remarks on property-identity above, may well be essential to an under-

standing of the traditional disputes over primary and secondary qualities.[71]

Let us return to the question of theoretical identification. Theoretical identities, according to the conception I advocate, are generally identities involving two rigid designators and therefore are examples of the necessary *a posteriori*. Now in spite of the arguments I gave before for the distinction between necessary and *a priori* truth, the notion of *a posteriori* necessary truth may still be somewhat puzzling. Someone may well be inclined to argue as follows: 'You have admitted that heat might have turned out not to have been molecular motion, and that gold might have turned out not to have been the element with the atomic number 79. For that matter, you also have acknowledged that Elizabeth II might have turned out not to be the daughter of George VI, or even to originate in the particular sperm and egg we had thought, and this table might have turned out to be made from ice made from water from the Thames. I gather that Hesperus might have turned out not to be Phosphorus. What then can you mean when you say that such eventualities are impossible? If Hesperus might have *turned out* not to be Phosphorus, then Hesperus might not have *been* Phosphorus. And similarly for the other cases: if the world could have *turned out* otherwise, it could have *been* otherwise. To deny this fact is to deny the self-evident modal principle that what is entailed by a possibility must itself be possible. Nor can you evade the difficulty by declaring the "might have" of "might have turned out otherwise" to be merely epistemic, in the way that "Fermat's Last Theorem might turn out to be true and might turn out to be false" merely expresses our present ignorance, and "Arithmetic might have turned out to be complete" signals our former ignorance. In these mathematical cases, we may have been ignorant, but it was in fact mathematically impossible for the answer to turn out other than it did. Not so in your favorite cases of essence and of identity between two rigid designators: it really is logically possible that gold should have turned out to be a compound, and this table might really have turned out not to be made of wood, let alone of a given particular block of wood. The contrast with the mathematical case could not be greater and would not be alleviated even if, as you suggest, there may be mathematical truths which it is impossible to know *a priori*.'

Perhaps anyone who has caught the spirit of my previous remarks can give my answer himself, but there is a clarification of my previous discussion which is relevant here. The objector is correct when he argues that if I hold that this table could not have been made of ice, then I must also hold that it could not have turned out to be made of ice; *it could have turned out that P*

[71] To understand this dispute, it is especially important to realize that yellowness is not a dispositional property, although it is related to a disposition. Many philosophers for want of any other theory of the meaning of the term 'yellow', have been inclined to regard it as expressing a dispositional property. At the same time, I suspect many have been bothered by the 'gut feeling' that yellowness is a manifest property, just as much 'right out there' as hardness or spherical shape. The proper account, on the present conception is, of course, that the reference of 'yellowness' is fixed by the description 'that (manifest) property of objects which causes them, under normal circumstances, to be seen as yellow (i.e., to be sensed by certain visual impressions)'; 'yellow', of course, does not *mean* 'tends to produce such and such a sensation'; if we had had different neutral structures, if atmospheric conditions had been different, if we had been blind, and so on, then yellow objects would have done no such thing. If one tries to revise the definition of 'yellow' to be, 'tends to produce such and such visual impressions under circumstances C', then one will find that the specification of the circumstances C either circularly involves yellowness or plainly makes the alleged definition into a scientific discovery rather than a synonymy. If we take the 'fixes a reference' view, then it is up to the physical scientist to identify the property so marked out in any more fundamental physical terms that he wishes.

Some philosophers have argued that such terms as 'sensation of yellow', 'sensation of heat', 'sensation of pain', and the like, could not be in the language unless they were identifiable in terms of external observable phenomena, such as heat, yellowness, and associated human behavior. I think that this question is independent of any view argued in the text.

entails that P could have been the case. What, then, does the intuition that the table might have turned out to have been made of ice or of anything else, that it might even have turned out not to be made of molecules, amount to? I think that it means simply that there might have been *a table* looking and feeling just like this one and placed in this very position in the room, which was in fact made of ice. In other words, I (or some conscious being) could have been *qualitatively in the same epistemic situation* that in fact obtains, I could have the same sensory evidence that I in fact have, about *a table* which was made of ice. The situation is thus akin to the one which inspired the counterpart theorists; when I speak of the possibility of the table turning out to be made of various things, I am speaking loosely. *This* table itself could not have had an origin different from the one it in fact had, but in a situation qualitatively identical to this one with respect to all the evidence I had in advance, the room could have contained *a table made of ice* in place of this one. Something like counterpart theory is thus applicable to the situation, but it applies only because we are *not* interested in what might have been true of *this particular* table, but in what might or might not be true of *a table* given certain evidence. It is precisely because it is *not* true that this table might have been made of ice from the Thames that we must turn here to qualitative descriptions and counterparts. To apply these notions to genuine *de re* modalities is, from the present standpoint, perverse.

The general answer to the objector can be stated, then, as follows: Any necessary truth, whether *a priori* or *a posteriori*, could not have turned out otherwise. In the case of some necessary *a posteriori* truths, however, we can say that under appropriate qualitatively identical evidential situations, an appropriate corresponding qualitative statement might have been false. The loose and inaccurate statement that gold might have turned out to be a compound should be replaced (roughly) by the statement that it is logically possible that there should have been a compound with all the properties originally known to hold of gold. The inaccurate statement that Hesperus might have turned out not to be Phosphorus should be replaced by the true contingency mentioned earlier in these lectures: two distinct bodies might have occupied, in the morning and the evening, respectively, the very positions actually occupied by Hesperus-Phosphorus-Venus.[72] The reason the example of Fermat's Last Theorem gives a different impression is that here no analogue suggests itself, except for the extremely general statement that, in the absence of proof or disproof, it is possible for *a mathematical conjecture* to be either true or false.

I have not given any general paradigm for the appropriate corresponding qualitative contingent statement. Since we are concerned with how things might have turned out otherwise, our general paradigm is to redescribe both the prior evidence and the statement qualitatively and claim that they are only contingently related. In the case of identities, using two rigid designators, such as the Hesperus-Phosphorus case above, there is a simpler paradigm which is often usable to at least approximately the same effect. Let 'R_1' and 'R_2' be the two rigid designators which flank the identity sign. Then '$R_1 = R_2$' is necessary if true. The references of 'R_1' and 'R_2', respectively, may well be fixed by nonrigid designators 'D_1' and 'D_2', in the Hesperus and Phosphorus cases these have the form 'the heavenly body in such-and-such position in the sky in the evening (morning)'. Then although '$R_1 = R_2$' is necessary,

[72] Some of the statements I myself make above may be loose and inaccurate in this sense. If I say, 'Gold *might* turn out not to be an element,' I speak correctly; 'might' here is *epistemic* and expresses the fact that the evidence does not justify *a priori* (Cartesian) certainty that gold is an element. I am also strictly correct when I say that the elementhood of gold was discovered *a posteriori*. If I say, 'Gold *might have* turned out not to be an element,' I seem to mean this metaphysically and my statement is subject to the correction noted in the text.

'$D_1 = D_2$' may well be contingent, and this is often what leads to the erroneous view that '$R_1 = R_2$' might have turned out otherwise.

I finally turn to an all too cursory discussion of the application of the foregoing considerations to the identity thesis. Identity theorists have been concerned with several distinct types of identifications: of a person with his body, of a particular sensation (or event or state of having the sensation) with a particular brain state (Jones's pain at 06 : 00 was his C-fiber stimulation at that time), and of *types* of mental states with the corresponding *types* of physical states (pain is the stimulation of C-fibers). Each of these, and other types of identifications in the literature, present analytical problems, rightly raised by Cartesian critics, which cannot be avoided by a simple appeal to an alleged confusion of synonymy with identity. I should mention that there is of course no obvious bar, at least (I say cautiously) none which should occur to any intelligent thinker on a first reflection just before bedtime, to advocacy of some identity theses while doubting or denying others. For example, some philosophers have accepted the identity of particular sensations with particular brain states while denying the possibility of identities between mental and physical *types*.[73] I will concern myself primarily with the type-type identities, and the philosophers in question will thus be immune to much of the discussion; but I will mention the other kinds of identities briefly.

Descartes, and others following him, argued that a person or mind is distinct from his body, since the mind could exist without the body. He might equally well have argued the same conclusion from the premise that the body could have existed without the mind.[74] Now the one response which I regard as plainly inadmissible is the response which cheerfully accepts the Cartesian premise while denying the Cartesian conclusion. Let 'Descartes' be a name, or rigid designator, of a certain person, and let '*B*' be a rigid designator of his body. Then if Descartes were indeed identical to *B*, the supposed identity, being an identity between two rigid designators, would be necessary, and Descartes could not exist without *B* and *B* could not exist without Descartes. The case is not at all comparable to the alleged analogue, the identity of the first Postmaster General with the inventor of bifocals. True, this identity obtains despite the fact that there could have been a first Postmaster General even though bifocals had never been invented. The reason is that 'the inventor of bifocals' is not a rigid designator; a world in which no one invented bifocals is not *ipso facto* a world in which Franklin did not exist. The alleged analogy therefore collapses; a philosopher who wishes

[73] Thomas Nagel and Donald Davidson are notable examples. Their views are very interesting, and I wish I could discuss them in further detail. It is doubtful that such philosophers wish to call themselves 'materialists'. Davidson, in particular, bases his case for his version of the identity theory on the supposed *impossibility* of correlating psychological properties with physical ones.

The argument against token-token identification in the text *does* apply to these views.

[74] Of course, the body *does* exist without the mind and presumably without the person, when the body is a corpse. This consideration, if accepted, would already show that a person and his body are distinct. (See David Wiggins, 'On Being at the Same Place at the Same Time', *Philosophical Review*, Vol. 77 (1968), pp. 90–5.) Similarly, it can be argued that a statue is not the hunk of matter of which it is composed. In the latter case, however, one might say instead that the former is 'nothing over and above' the latter; and the same device might be tried for the relation of the person and the body. The difficulties in the text would not then arise in the same form, but analogous difficulties would appear. A theory that a person is nothing over and above his body in the way that a statue is nothing over and above the matter of which it is composed, would have to hold that (necessarily) a person exists if and only if his body exists and has a certain additional physical organization. Such a thesis would be subject to modal difficulties similar to those besetting the ordinary identity thesis, and the same would apply to suggested analogues replacing the identification of mental states with physical states. A further discussion of this matter must be left for another place. Another view which I will not discuss, although I have little tendency to accept it and am not even certain that it has been set out with genuine clarity, is the so-called functional state view of psychological concepts.

SELECTIONS from KRIPKE, NAMING & NECESSITY, PAGE 10

to refute the Cartesian conclusion must refute the Cartesian premise, and the latter task is not trivial.

Let '*A*' name a particular pain sensation, and let '*B*' name the corresponding brain state, or the brain state some identity theorist wishes to identify with *A*. *Prima facie*, it would seem that it is at least logically possible that *B* should have existed (Jones's brain could have been in exactly that state at the time in question) without Jones feeling any pain at all, and thus without the presence of *A*. Once again, the identity theorist cannot admit the possibility cheerfully and proceed from there; consistency, and the principle of the necessity of identities using rigid designators, disallows any such course. If *A* and *B* were identical, the identity would have to be necessary. The difficulty can hardly be evaded by arguing that although *B* could not exist without *A*, *being a pain* is merely a contingent property of *A*, and that therefore the presence of *B* without pain does not imply the presence of *B* without *A*. Can any case of essence be more obvious than the fact that *being a pain* is a necessary property of each pain? The identity theorist who wishes to adopt the strategy in question must even argue that *being a sensation* is a contingent property of *A*, for *prima facie* it would seem logically possible that *B* could exist without any sensation with which it might plausibly be identified. Consider a particular pain, or other sensation, that you once had. Do you find it at all plausible that *that very sensation* could have existed without being a sensation, the way a certain inventor (Franklin) could have existed without being an inventor?

I mention this strategy because it seems to me to be adopted by a large number of identity theorists. These theorists, believing as they do that the supposed identity of a brain state with the corresponding mental state is to be analyzed on the paradigm of the contingent identity of Benjamin Franklin with the inventor of bifocals, realize that just as his contingent activity made Benjamin Franklin into the inventor of bifocals, so some contingent property of the brain state must make it into a pain. Generally they wish this property to be one statable in physical or at least 'topic-neutral' language, so that the materialist cannot be accused of positing irreducible non-physical properties. A typical view is that *being a pain*, as a property of a physical state, is to be analyzed in terms of the 'causal role' of the state,[75] in terms of the characteristic stimuli (e.g., pinpricks) which cause it and the characteristic behavior it causes. I will not go into the details of such analyses, even though I usually find them faulty on specific grounds in addition to the general modal considerations I argue here. All I need to observe here is that the 'causal role' of the physical state is regarded by the theorists in question as a contingent property of the state, and thus it is supposed to be a contingent property of the state that it is a mental state at all, let alone that it is something as specific as a pain. To repeat, this notion seems to me self-evidently absurd. It amounts to the view that the *very pain I now have* could have existed without being a mental state at all.

I have not discussed the converse problem, which is closer to the original Cartesian consideration—namely, that just as it seems that the brain state could have existed without any pain, so it seems that the pain could have existed without the corresponding brain state. Note that *being a brain state* is evidently an essential property of *B* (the brain state). Indeed, even more is true: not only being a brain state, but even being a brain state of a specific type is an essential property of *B*. The configuration of brain cells whose presence at a given time constitutes the presence of *B* at that time is essential to *B*, and in its absence *B* would not have existed. Thus someone who

[75] For example, David Armstrong, *A Materialist Theory of the Mind*, London and New York, 1968, see the discussion review by Thomas Nagel, *Philosophical Review* **79** (1970), pp. 394–403; and David Lewis, 'An Argument for the Identity Theory', *The Journal of Philosophy*, pp. 17–25.

SELECTIONS from KRIPKE, NAMING & NECESSITY, PAGE 11

wishes to claim that the brain state and the pain are identical must argue that the pain A could not have existed without a quite specific type of configuration of molecules. If $A = B$, then the identity of A with B is necessary, and any essential property of one must be an essential property of the other. Someone who wishes to maintain an identity thesis cannot simply *accept* the Cartesian intuitions that A can exist without B, that B can exist without A, that the correlative presence of anything with mental properties is merely contingent to B, and that the correlative presence of any specific physical properties is merely contingent to A. He must explain these intuitions away, showing how they are illusory. This task may not be impossible; we have seen above how some things which appear to be contingent turn out, on closer examination, to be necessary. The task, however, is obviously not child's play, and we shall see below how difficult it is.

The final kind of identity, the one which I said would get the closest attention, is the type-type sort of identity exemplified by the identification of pain with the stimulation of C-fibers. These identifications are supposed to be analogous with such scientific type-type identifications as the identity of heat with molecular motion, of water with hydrogen hydroxide, and the like. Let us consider, as an example, the analogy supposed to hold between the materialist identification and that of heat with molecular motion; both identifications identify two types of phenomena. The usual view holds that the identification of heat with molecular motion and of pain with the stimulation of C-fibers are both contingent. We have seen above that since 'heat' and 'molecular motion' are both rigid designators, the identification of the phenomena they name is necessary. What about 'pain' and 'C-fiber stimulation'? It should be clear from the previous discussion that 'pain' is a rigid designator of the type, or phenomenon, it designates: if something is a pain it is essentially so, and it seems absurd to suppose that pain could have been some phenomenon other than the one it is. The same holds for the term 'C-fiber stimulation', provided that 'C-fibers' is a rigid designator, as I will suppose here. (The supposition is somewhat risky, since I know virtually nothing about C-fibers, except that the stimulation of them is said to be correlated with pain.[76] The point is unimportant; if 'C-fibers' is not a rigid designator, simply replace it by one which is, or suppose it used as a rigid designator in the present context.) Thus the identity of pain with the stimulation of C-fibers, if true, must be *necessary*.

So far the analogy between the identification of heat with molecular motion and pain with the stimulation of C-fibers has not failed; it has merely turned out to be the opposite of what is usually thought—both, if true, must be necessary. This means that the identity theorist is committed to the view that there could not be a C-fiber stimulation which was not a pain nor a pain which was not a C-fiber stimulation. These consequences are certainly surprising and counterintuitive, but let us not dismiss the identity theorist too quickly. Can he perhaps show that the apparent possibility of pain not having turned out to be C-fiber stimulation, or of there being an instance of one of

[76] I have been surprised to find that at least one able listener took my use of such terms as 'correlated with', 'corresponding to', and the like as already begging the question against the identity thesis. The identity thesis, so he said, is not the thesis that pains and brain states are correlated, but rather that they are identical. Thus my entire discussion presupposes the anti-materialist position that I set out to prove. Although I was surprised to hear an objection which concedes so little intelligence to the argument, I have tried especially to avoid the term 'correlated' which seems to give rise to the objection. Nevertheless, to obviate misunderstanding, I shall explain my usage. Assuming, at least *arguendo*, that scientific discoveries have turned out so as not to refute materialism from the beginning, both the dualist and the identity theorist agree that there is a correlation or correspondence between mental states and physical states. The dualist holds that the 'correlation' relation in question is irreflexive; the identity theorist holds that it is simply a special case of the identity relation. Such terms as 'correlation' and 'correspondence' can be used neutrally without prejudging which side is correct.

the phenomena which is not an instance of the other, is an illusion of the same sort as the illusion that water might not have been hydrogen hydroxide, or that heat might not have been molecular motion? If so, he will have rebutted the Cartesian, not, as in the conventional analysis, by accepting his premise while exposing the fallacy of his argument, but rather by the reverse—while the Cartesian argument, given its premise of the contingency of the identification, is granted to yield its conclusion, the premise is to be exposed as superficially plausible but false.

Now I do not think it likely that the identity theorist will succeed in such an endeavor. I want to argue that, at least, the case cannot be interpreted as analogous to that of scientific identification of the usual sort, as exemplified by the identity of heat and molecular motion. What was the strategy used above to handle the apparent contingency of certain cases of the necessary *a posteriori*? The strategy was to argue that although the statement itself is necessary, someone could, *qualitatively* speaking, be in the same epistemic situation as the original, and in such a situation a *qualitatively* analogous statement could be false. In the case of identities between two rigid designators, the strategy can be approximated by a simpler one: Consider how the references of the designators are determined; if these coincide only contingently, it is this fact which gives the original statement its illusion of contingency. In the case of heat and molecular motion, the way these two paradigms work out is simple. When someone says, inaccurately, that heat might have turned out not to be molecular motion, what is true in what he says is that someone could have sensed a phenomenon in the same way we sense heat, that is, feels it by means of its production of the sensation we call 'the sensation of heat' (call it '*S*'), even though that phenomenon was not molecular motion. He means, additionally, that the planet might have been inhabited by creatures who did not get *S* when they were in the presence of molecular motion, though perhaps getting it in the presence of something else. Such creatures would be, in some qualitative sense, in the same epistemic situation as we are, they could use a rigid designator for the phenomenon that causes sensation *S* in them (the rigid designator could even be 'heat'), yet it would not be molecular motion (and therefore not heat!), which was causing the sensation.

Now can something be said analogously to explain away the feeling that the identity of pain and the stimulation of C-fibers, if it is a scientific discovery, could have turned out otherwise? I do not see that such an analogy is possible. In the case of the apparent possibility that molecular motion might have existed in the absence of heat, what seemed really possible is that molecular motion should have existed without being *felt as heat*, that is, it might have existed without producing the sensation *S*, the sensation of heat. In the appropriate sentient beings is it analogously possible that a stimulation of C-fibers should have existed without being felt as pain? If this is possible, then the stimulation of C-fibers can itself exist without pain, since for it to exist without being *felt as pain* is for it to exist without there *being any* pain. Such a situation would be in flat out contradiction with the supposed necessary identity of pain and the corresponding physical state, and the analogue holds for any physical state which might be identified with a corresponding mental state. The trouble is that the identity theorist does not hold that the physical state merely *produces* the mental state, rather he wishes the two to be identical and thus *a fortiori* necessarily co-occurrent. In the case of molecular motion and heat there is something, namely, the sensation of heat, which is an intermediary between the external phenomenon and the observer. In the mental-physical case no such intermediary is possible, since here the physical phenomenon is supposed to be identical with the

internal phenomenon itself. Someone can be in the same epistemic situation as he would be if there were heat, even in the absence of heat, simply by feeling the sensation of heat; and even in the presence of heat, he can have the same evidence as he would have in the absence of heat simply by lacking the sensation *S*. No such possibility exists in the case of pain and other mental phenomena. To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain; to be in the same epistemic situation that would obtain in the absence of a pain *is* not to have a pain. The apparent contingency of the connection between the mental state and the corresponding brain state thus cannot be explained by some sort of qualitative analogue as in the case of heat.

We have just analyzed the situation in terms of the notion of a qualitatively identical epistemic situation. The trouble is that the notion of an epistemic situation qualitatively identical to one in which the observer had a sensation *S* simply *is* one in which the observer had that sensation. The same point can be made in terms of the notion of what picks out the reference of a rigid designator. In the case of the identity of heat with molecular motion the important consideration was that although 'heat' is a rigid designator, the reference of that designator was determined by an accidental property of the referent, namely the property of producing in us the sensation *S*. It is thus possible that a phenomenon should have been rigidly designated in the same way as a phenomenon of heat, with its reference also picked out by means of the sensation *S*, without that phenomenon being heat and therefore without its being molecular motion. Pain, on the other hand, is not picked out by one of its accidental properties; rather it is picked out by the property of being pain itself, by its immediate phenomenological quality. Thus pain, unlike heat, is not only rigidly designated by 'pain' but the reference of the designator is determined by an essential property of the referent. Thus it is not possible to say that although pain is necessarily identical with a certain physical state, a certain phenomenon can be picked out in the same way we pick out pain without being correlated with that physical state. If any phenomenon is picked out in exactly the same way that we pick out pain, then that phenomenon *is* pain.

Perhaps the same point can be made more vivid without such specific reference to the technical apparatus in these lectures. Suppose we imagine God creating the world; what does He need to do to make the identity of heat and molecular motion obtain? Here it would seem that all He needs to do is to create the heat, that is, the molecular motion itself. If the air molecules on this earth are sufficiently agitated, if there is a burning fire, then the earth will be hot even if there are no observers to see it. God created light (and thus created streams of photons, according to present scientific doctrine) before He created human and animal observers; and the same presumably holds for heat. How then does it appear to us that the identity of molecular motion with heat is a substantive scientific fact, that the mere creation of molecular motion still leaves God with the additional task of making molecular motion into heat? This feeling is indeed illusory, but what *is* a substantive task for the Deity is the task of making molecular motion felt as heat. To do this He must create some sentient beings to insure that the molecular motion produces the sensation *S* in them. Only after he has done this will there be beings who can learn that the sentence 'Heat is the motion of molecules' expresses an *a posteriori* truth in precisely the same way that we do.

What about the case of the stimulation of C-fibers? To create this phenomenon, it would seem that God need only create beings with C-fibers capable of the appropriate type of physical stimulation; whether the beings are conscious or not is irrelevant here. It would seem, though, that to make the C-fiber stimulation correspond to pain, or be felt as pain, God must

do something in addition to the mere creation of the C-fiber stimulation; He must let the creatures feel the C-fiber stimulation as *pain*, and not as a tickle, or as warmth, or as nothing, as apparently would also have been within His powers. If these things in fact are within His powers, the relation between the pain God creates and the stimulation of C-fibers cannot be identity. For if so, the stimulation could exist without the pain; and since 'pain' and 'C-fiber stimulation' are rigid, this fact implies that the relation between the two phenomena is not that of identity. God had to do some work, in addition to making the man himself, to make a certain man be the inventor of bifocals; the man could well exist without inventing any such thing. The same cannot be said for pain; if the phenomenon exists at all, no further work should be required to make it into pain.

In sum, the correspondence between a brain state and a mental state seems to have a certain obvious element of contingency. We have seen that identity is not a relation which can hold contingently between objects. Therefore, if the identity thesis were correct, the element of contingency would not lie in the relation between the mental and physical states. It cannot lie, as in the case of heat and molecular motion, in the relation between the phenomenon (= heat = molecular motion) and the way it is felt or appears (sensation *S*), since in the case of mental phenomena there is no 'appearance' beyond the mental phenomenon itself.

Here I have been emphasizing the possibility, or apparent possibility, of a physical state without the corresponding mental state. The reverse possibility, the mental state (pain) without the physical state (C-fiber stimulation) also presents problems for the identity theorists which cannot be resolved by appeal to the analogy of heat and molecular motion.

I have discussed similar problems more briefly for views equating the self with the body, and particular mental events with particular physical events, without discussing possible countermoves in the same detail as in the type-type case. Suffice it to say that I suspect that the considerations given indicate that the theorist who wishes to identify various particular mental and physical events will have to face problems fairly similar to those of the type-type theorist; he too will be unable to appeal to the standard alleged analogues.

That the usual moves and analogies are not available to solve the problems of the identity theorist is, of course, no proof that no moves are available. I certainly cannot discuss all the possibilities here. I suspect, however, that the present considerations tell heavily against the usual forms of materialism. Materialism, I think, must hold that a physical description of the world is a *complete* description of it, that any mental facts are 'ontologically dependent' on physical facts in the straightforward sense of following from them by necessity. No identity theorist seems to me to have made a convincing argument against the intuitive view that this is not the case.[77]

[77] Having expressed these doubts about the identity theory in the text, I should emphasize two things: first, identity theorists have presented positive arguments for their view, which I certainly have not answered here. Some of these arguments seem to me to be weak or based on ideological prejudices, but others strike me as highly compelling arguments which I am at present unable to answer convincingly. Second, rejection of the identity thesis does not imply acceptance of Cartesian dualism. In fact, my view above that a person could not have come from a different sperm and egg from the ones from which he actually originated implicitly suggests a rejection of the Cartesian picture. If we had a clear idea of the soul or the mind as an independent, susbsistent, spiritual entity, why should it have to have any necessary connection with particular material objects such as a particular sperm or a particular egg? A convinced dualist may think that my views on sperms and eggs beg the question against Descartes. I would tend to argue the other way; the fact that it is hard to imagine me coming from a sperm and egg different from my actual origins seems to me to indicate that we have no such clear conception of a soul or self. In any event, Descartes' notion seems to have been rendered dubious ever since Hume's critique of the notion of a Cartesian self. I regard the mind-body problem as wide open and extremely confusing.