Kripke's *Naming and Necessity*: Lecture III

PHIL 83104
October 25, 2011

Kripke ends Lecture II, and starts Lecture III, by discussing certain sorts of identity sentences as examples of the necessary a posteriori. Lecture III goes on to focus on two important further classes of such claims: attributions of essential properties to things, and theoretical identity claims. Ultimately, Kripke thinks, attention to claims of this sort will shed light on the mind/body problem, which he discusses briefly at the end of the lecture.

1. Attributions of essential properties & the necessary a posteriori (106-116)

How could essential properties generate examples of the necessary a posteriori? Let n be a name, and F be a predicate which expresses an essential property of the referent of n. Then the above explanation of the essential/accidental distinction is enough to show that ⌜n is F⌝ will express a necessary truth. But the proposition expressed by this sentence might well be a posteriori as well, since it might take empirical investigation to find out whether the referent of 'n' in fact has the property expressed by 'F'. As Kripke puts it:

> "...other considerations ...about an object having essential properties, can only be regarded correctly, in my view, if we recognize the distinction between a prioricity and necessity. One might very well discover essence empirically." (110)

We can approach a similar point another way (this is discussed by Kripke in his paper, "Identity and Necessity"). It might be the case that for some property, we can know a priori that, if some object has that property, it has that property essentially. For example, it might be the case that I know of each of you that, if you are human, you are necessarily (essentially) human. But it might take empirical work to determine that you are in fact human, rather than a cleverly disguised robot. In this case, we will have a necessary and a priori claim combining with a contingent and a posteriori claim to yield an example of the contingent a posteriori.

Consider, for example, the following argument:

1. The object before me is a human being.
2. o is the object before me.
3. o is a human being. (1,2)
4. $\forall$x(x is a human being$\rightarrow \Box$(x is a human being))

C. Necessarily, o is a human being. (3,4)

Premises (1) and (2) are ordinary contingent claims that are, presumably, knowable only a posteriori. But if one knows (1) and (2), one is then in a position to deduce (3), which is an instance of the necessary a posteriori (ignoring for now issues about o's possible nonexistence). But you might still wonder how we could know that (3) is a necessary truth; how could we ever know, or some a posteriori proposition, that it is necessary? Well, we might begin with a claim like (4), which seems like a necessary and a priori truth. Once we know (3) and (4), we are in a position to deduce (C), which says of our necessary a posteriori proposition (3) that it is necessary.

This argument thus shows how, on the basis of knowledge of necessary a priori truths and contingent a posteriori ones, we can come to know that a certain a posteriori truth is necessary.

Kripke goes on to discuss two ways of generating true essentialist claims.

1.1. Essentiality of origins (111-114)

Kripke suggests as a plausible essentialist principle the view that if a material object has its origin in a certain bit of matter, then it could not have existed without having that origin. He discusses the case of a particular table:

> "In the case of this table, we may not know what block of wood the table came from. Now could this table have been made from a completely different block of wood, or even of water cleverly hardened into ice — water taken from the Thames River? We could conceivably discover that, contrary to what we now think, this table is indeed made of ice from the river. But

> suppose that it is not. Then, though we can imagine making a table out of another block of wood or even from ice, identical in appearance with this one, and though we could have put it in this very position in the room, it seems to me that this is not to imagine this table as made of wood or ice, but rather it is to imagine another table, resembling this one in all external details, made of another block of wood, or even of ice." (113-114)

1.2. Essentiality of constitution (114 n. 57, 126-127)

Kripke defends a similar principle about the material constitution of things. Could something composed of molecules have existed — could it have been that thing — without being composed of molecules?

1.3. More trivial examples

There's been much discussion over the question of whether essentialist claims of these sorts are true. Here it is less important to focus on this question than to see how these claims, if true, would be instances of the necessary a posteriori.

You might also think that there are other, less interesting examples of plausible essentialist claims. Consider, e.g. the following:

> Saul Kripke is essentially not a fried egg.
> I am essentially non-identical to Saul Kripke.
>
> 4 is essentially greater than 3.
> ∅ essentially lacks members.

The first two of these can be used to generate examples of the necessary a posteriori, using the formula sketched above.

2. Theoretical identities (116-143)

Kripke thinks that theoretical identities of the sort at which science aims are a third source of necessary a posteriori truths. This view is connected with certain theses about natural kind terms, and their similarity to proper names.

2.1. Analogies between proper names and natural kind terms

2.1.1. Natural kind terms and descriptions

The first similarity between names and natural kind terms comes in their relations to various associated definite descriptions. Just as names are associated with various definite descriptions — e.g., 'Aristotle' with 'the greatest philosopher of antiquity', 'the teacher of Alexander', etc. — so natural kind terms are associated with various descriptions — among the examples Kripke discusses are 'gold' and 'the yellow metal', 'heat' and 'the

cause of sensation S', 'water' and 'the clear drinkable liquid', 'tiger' and 'the striped quadrupedal carnivorous feline.'

Just as Kripke argued that names are not synonymous with the descriptions associated with them, so he argues that natural kind terms are not synonymous with the descriptions associated with them. This is the point of the discussion of Kant's idea that 'Gold is a yellow metal' is analytic, and that we can imagine seeing a three-legged tiger. Kripke concludes that the role that such descriptions play in the case of natural kind terms is analogous to the role that they play in the case of names. The idea there was that a description might be used initially to fix the reference of a name, but typically need not be known by users later in the history of the name. Kripke thinks the same about natural kind terms. We might have introduced the natural kind term 'water' via the description 'the clear liquid over there', but this neither gives the meaning of 'water' nor need be recognized by later speakers who understand the term.

2.1.2. Natural kind terms and rigid designation

Kripke thinks that the similarities between names and kind terms extend beyond their 'nondescriptionality.' He further thinks that natural kind terms, like names, are rigid designators.

One of Kripke's aims is to show that theoretical identity sentences are necessary, if true. Recall that he established a similar thesis about identity sentences involving names. Because ordinary proper names are rigid designators, he said, any true identity sentence of the form 'n is m' will be a necessary truth.

It is clear that he thinks that we can give a similar explanation of the necessity of theoretical identities. He says,

> "Theoretical identities, according to the conception I advocate, are generally identities involving two rigid designators and therefore are examples of the necessary a posteriori." (140)

Among the theoretical identity statements which Kripke thinks to be necessary for this reason are:

> Water is H2O.
>
> Lightning is electricity.
>
> Heat is molecular motion.
>
> Gold is the element with atomic number 79.
>
> Cats are animals.

We have an understanding of why identity sentences involving two coreferential rigidly designating singular terms are necessary if true. What we now have to ask is: how does this explanation carry over to the case of theoretical identities? There are two reasons to be skeptical that it does: (1) these 'theoretical identities' do not appear to be identity statements, and (2) we have no grip yet on what it means for a general term like 'water' to be a rigid designator.

2.1.3. Are theoretical identities really identity statements?

In the case of identity sentences like 'Hesperus is Phosphorus' we have a claim that the object which is the referent of 'Hesperus' stands in the identity relation to the object which is the referent of 'Phosphorus.' If we were to apply this paradigm to, e.g., 'Water is H2O', this would mean that the claim is true iff the referent of 'water' stands in the identity relation to the referent of 'H2O.'

Supposing that the referents of these general terms are the sets of objects to which they apply, an obvious problem for this view is that it does not apply to some of the above examples; for example, not all electricity is lightning, so the referents of 'lightning' and 'electricity' will not stand in the identity relation. The same goes for 'Cats are animals.'

An alternate reading of the logical forms of theoretical identities: they are not identity statements, but universally quantified conditionals and biconditionals. E.g.:

$\forall$x (x is water $\equiv$ x is H2O)
$\forall$x (x is a cat $\rightarrow$ x is an animal)

2.1.4. What would it mean for natural kind terms to be rigid designators?

Still, you might think, we can give an explanation of why these conditionals and biconditionals are necessary if true using the notion of rigid designation. But here we run into further problems; Kripke never really explains what it would mean for a general term, as opposed to a proper name, to be a rigid designator.

Consider the following possible interpretations of the claim that natural kind terms are rigid designators, and the reasons why they are unsuccessful:

> A predicate is a rigid designator iff it has the same reference with respect to every possible world.

On this characterization, virtually no predicates, and none of Kripke's examples, will be rigid. Consider 'cat'. Surely there is a possible world w in which there are no cats; but then there is a possible world with respect to which 'cat' has a different referent than it does with respect to the actual world.

> A predicate is a rigid designator iff if the predicate applies to an object in at least one possible world, it applies to that object with respect to every possible world.

Though this is a plausible-sounding extension of the idea of rigidity for singular terms, it fails to explain why statements like the above are necessary if true. It does not rule out, for example, there being a possible world w at which something is a cat but not an animal — so long as that thing is essentially a cat, and essentially a non-animal.

It is an open question whether there is any understanding of what it means for a predicate to be a rigid designator which both fits well with Kripke's text and explains why he thinks that theoretical identities of the sort listed above are necessary, if true. For further discussion, see Soames's *Beyond Rigidity*.

2.2. A neo-Kripkean explanation of the necessity of theoretical identities

The seeming failure of Kripke's explanation of the modal status of theoretical identities does not decide the question of whether Kripke's claim that such sentences are necessary if true is correct. Questions about rigid designation aside, his claim that these claims are necessary has some plausibility. Consider what he says about 'Gold is the element with atomic number 79':

> "Gold apparently has the atomic number 79. Is it a necessary or a contingent property of gold that it has the atomic number 79? . . . Suppose we now find some other yellow metal, or some other yellow thing, with all the properties by which we originally identified gold, and many of the additional ones that we have discovered later. An example of one with many of the initial properties is iron pyrites, 'fool's gold.' As I have said, we wouldn't say that this substance is gold. So far we are speaking of the actual world. Now consider a possible world. Consider a counterfactual situation in which, let us say, fool's gold or iron pyrites was actually found in various mountains of the United States, or in areas of South Africa and the Soviet Union. Suppose that all the areas which actually contain gold now, contained iron pyrites instead, or some other substance which counterfeited the superficial properties of gold but lacked its atomic structure. Would we say, of this counterfactual situation, that in that situation gold would not have been an element (because pyrites is not an element)? It seems to me that we would not. We would instead describe this as a situation in which a substance, say iron pyrites, which is not gold, would have been found in the very mountains which actually contain gold and would have had the very properties by which we commonly identify gold. But it would not be gold; it would be something else. ...(Once again, whether people counterfactually would have called it 'gold' is irrelevant. . . .) . . . Given that gold is this element, any other

> substance, even though it looks like gold and is found in the very places where we in fact find gold, would not be gold." (123-125)

Similar arguments can be given in the case of other theoretical identities. But what *makes* these conditional and biconditional claims necessary, if not that natural kind terms are rigid designators?

Here are two (compatible) answers to this question:

1. One view is that this is not so much a separate class of necessary a posteriori claims as a sort of special case of essentialist claims. Consider, for example, "All whales are mammals." Perhaps it is a necessary, a priori truth that if every member of a certain genus is also a member of a certain class, then it is essential to that genus that every member of it be a member of that class. It could still be a posteriori that every whale is in fact a mammal; but these two claims would combine — in a way familiar from our discussion of essentialist claims — to give us the result that the claim that every whale is a mammal is an example of the necessary a posteriori.

2. A second explanation of the modal character of these claims comes from Kripke's ideas about how some natural kind terms like "water" are introduced. The rough idea is that such terms are introduced by reference to a sample, and designate all things which are of the same kind as the sample. We might, of course, mean different things by "same kind" — we might be referring to a species kind, a chemical kind, etc. Let's suppose that for a given type of kind, a thing can belong to at most one kind of that type. (For example, a thing can belong to at most one chemical kind, let's say.) So suppose that "water" designates a chemical kind, and that we discover that the things in the original sample are of chemical kind H2O. It follows that the kind water=the kind H2O — which, given the necessity of identity, implies that it is a necessary truth that if something is water, then it is H2O.

So, even if we can't derive the conclusion that these sorts of claims are typically necessary a posteriori from Kripke's discussion of rigid designation, Kripke's claims about these theoretical identities still seem quite plausible.

2.3. The illusion of contingency

However, this seems to run into the following objection: it seems clear that heat might *have turned out not to be* molecular motion, and that gold might *have turned out not to be* the element with atomic number 79. But when we say this, we seem to be affirming the possibility of something: namely, the possibility that heat is not molecular motion, and the possibility that gold is not the element with atomic number 79. But the possibility of these claims being true straightforwardly conflicts with Kripke's claim that the above theoretical identities are necessarily true, since 'Necessarily p' and 'Possibly not-p' are inconsistent. (See pp. 141 ff for Kripke's discussion of this.) What's going on?

Kripke thinks that this appearance that theoretical identities are contingent is an illusion, and his aim is to to explain it away:

> "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." (142-143)

Why this is a plausible re-description of the intuition behind 'Gold might not have turned out to have atomic number 79'; what we are imagining when we are imagining gold turning out to have a different atomic number.

It is important to be clear about why this explanation works: we are supposing that we have in mind a situation which is like the actual world in respect of how things appear to us, but different in respect of how things are. What we will see later is that there are special reasons why this explanation will not work in the case of mind-body identities; and this is crucial to Kripke's argument that the mind is not identical to the body.

3. Materialism & the mind/body problem (144-)

Kripke next turns his attention to the mind-body problem. The discussion here brings to bear many of the results from earlier in the book, including rigid designation, the necessity of identity, the possibility of necessary a posteriori truths, and explanations of the illusion that certain a posteriori necessary truths are contingent.

3.1. Three identity theories

An identity theory of some mental phenomenon is a theory which says that that phenomenon (state, property) is identical to some physical phenomenon. Kripke argues that no identity theory can be correct. In doing so, he distinguishes between three different identity theories:

> "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 6:00 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)." (144)

These are not so much three varieties of identity theory as they are three different topics which one can be an identity theorist about: substances which have mental properties, the mental properties themselves, and instances, or tokens, of those mental properties. In each case, we have an opposition between identity theories and dualist theories (along with, perhaps, some intermediate positions): substance identity theory/dualism, property (type) identity theory/dualism, and token identity theory/dualism.

3.2. Three arguments against identity theories

Kripke presents arguments against all three sorts of identity theories. The arguments themselves, as Kripke says, have been around at least since Descartes; the main original contribution of Kripke's discussion is to show that certain apparently convincing replies to the Cartesian arguments are unsuccessful.

The form of the Cartesian argument is to point out the possibility of some mental phenomenon without the physical phenomenon alleged to be identical to it by the identity theorist, and to conclude from there that the mental and physical phenomena are actually not identical. The three versions of this argument which Kripke gives are:

Argument against person-body identity theories:

Possibly, this person/= this body. (Relevant scenarios: I exist without this body, or this body exists without me.)

This person $\neq$ this body.

Argument against token identity theories:

(Following Kripke (p. 146), let 'A' be the name of a particular pain sensation, and let 'B' name the brain state with which A is claimed to be identical. Think of 'A' as meaning 'this pain.')

Possibly, A$\neq$ B. (Relevant scenarios: this very pain sensation exists without this very brain state, or this brain state exists without this pain sensation.)

A$\neq$B.

Argument against type (property) identity theories:

Possibly, pain $\neq$ C-fiber firing. (Relevant scenarios: some creatures experience pain without their C-fibers firing, or some creatures have C-fibers which fire but no pain.)

Pain $\neq$ C-fiber firing.

In each case, the premise is that some non-identity is possibly true. We establish this by considering a possible world in which one of the terms of the identity, but not the other, does not refer. (These possible worlds are the 'relevant scenarios' listed alongside the arguments.) The identity statement in question is thus false with respect to this possible scenario, and hence possibly false. This is what each of the premises of the Cartesian arguments say. We then conclude from the fact that the identity possibly fails to hold that it actually fails to hold. (More on this transition below.)

3.3. Materialist responses

Faced with these arguments, the materialist has two options: claim that the argument is invalid, or deny the premise. The material earlier in *Naming and Necessity* provides arguments which count against either move.

3.3.1. Contingent identities

Suppose that we take the first option, and deny that the arguments are valid. Then we must deny that the relevant identities are necessary if true since, if they were necessary if true, their possible falsehood would show their falsehood *simpliciter*. But don't we come across contingent identities all the time — for example, 'The first Postmaster General of the United States = the inventor of bifocals'? Given examples like this, it might well seem that the Cartesian owes the materialist an explanation of why the identities posited by the materialist cannot just be contingent identities.

But recall our discussion of Kripke's treatment of identity sentences. There we were given a convincing argument for the necessity of identity, and an explanation of apparent contingent identities as involving non-rigid designators. So, if this response to the Cartesian is to be convincing, it must be the case that some of the terms which figure in the relevant identities are non-rigid. Are they?

Consider the first two arguments, since the third raises its own difficulties. It seems that none of the four terms which figure in these two identity claims are non-rigid; 'this person' and 'this body' seem to rigidly designate me and my body, respectively, and 'A' and 'B' were supposed to be introduced as names for particular sensations and particular brain states/events. So it is plausible that each of these four terms is a rigid designator. If so, each of the identities will be necessary if true, and this is enough for the Cartesian arguments to be valid.

But it is important to see that there is a sense in which it doesn't matter much whether the particular terms used in these arguments are rigid designators; for, if they are not, we could replace them with rigid designators. All the Cartesian needs to run his arguments against the identity theorist is that there be some rigid designators or other which refer to

the relevant substances and events/states. Accordingly, the anti-Cartesian proponent of contingent identities must show that we rigidly designate the items in question. But it is difficult to see why this should be so.

Consider how the third argument, having to do with mental properties. This presents extra complications because the class of statements to which Kripke assimilates the view of type identity theorist presents extra complications:

> "The final kind of identity . . . 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 identification of heat with molecular motion, of water with hydrogen hydroxide, and the like." (148)

Kripke thinks, as we have seen, that these scientific identifications are necessary if true; so he thinks that the claim of the type identity theorist is also necessary if true.

Here we run into a problem with the evaluation of Kripke's argument. In our discussion of theoretical identities, we found Kripke's claim that theoretical identifications are necessary if true to be plausible, but found his argument for this claim to be unconvincing. The problems with the argument were that (i) it rests on an under-specified extension of the term 'rigid designator' from singular terms to predicates, and (ii) it rests on treating theoretical identities as identity sentences, whereas in many cases ('Cats are animals') it is clear that they cannot be treated in this way, but must rather be treated as universally quantified conditionals and biconditionals.

One way around this problem is as follows. Even though we cannot treat all of the statements with which Kripke was concerned as identity statements, we *can* treat the claim of the type identity theorist as an identity statement. It is simply an identity involving a property rather than one involving an object. So, in this context, 'pain' is to be thought of as functioning not as a predicate, but as a singular term which refers to a property. Construed this way, the claim that 'pain' is a rigid designator is perfectly well defined: it is a rigid designator just in case it refers to the same property with respect to every possible world. Same goes for 'C-fiber firing.' Then we can block the present reply to the Cartesian argument as above.

3.3.2. A posteriori necessities and the illusion of contingency

So we have headed off one line of reply to the Cartesian arguments: the claim that the arguments are invalid. There's only one other possible reply: that the arguments are valid, but unsound. Since the arguments have only one premise, the only way to go here is to argue that the claims about possibility employed as the premises in these arguments are false. But, since these premises would clearly be true if the 'relevant scenarios' listed

above were possible, this reply amounts to the claim that these seemingly possible scenarios are not really possible.

Why, without argument, this sounds implausible; a link between conceivability and possibility, and the conceivability of the relevant scenarios.

The anti-Cartesian might back up his claim by using Kripke's own work. After all, Kripke has shown that there can be a posteriori necessities, and has shown that in the case of a posteriori necessities there will usually be an illusion that certain claims are possible which are not, in fact, possible. But we can explain this intuition away, as Kripke has shown us. Let's recall how this works.

Let's focus on the claim that gold is the element with atomic number 79. Initially, this seems contingent, since we can imagine, as we put it, 'it turning out that gold has a different atomic number.' But if we can imagine this, can't we clearly conceive of it's being the case that gold does not have atomic number 79? Kripke thinks not. He thinks that what we imagine when we imagine it's turning out that gold does not have atomic number 79 is a situation qualitatively identical to our actual one, in which someone discovers that some stuff with has the same superficial identifying marks as gold is a compound rather than an element. But, if we think carefully about this situation, we will see that we are not really inclined to say that this is a scenario in which gold is a compound. Rather, we are inclined to say that it is a scenario in which the stuff which looks like gold is not really gold. Thus we have a proposition which seemed clearly to be possible — that gold is a compound rather than an element — which turns out on closer inspection not to be possible.

Now it seems that we have given the anti-Cartesian all that he could want. We have explained that certain initial intuitions about possibility can be false, and have even offered a plausible explanation of why these illusions of contingency arise. Can't the anti-Cartesian just say, at this point, that our intuitions that I could exist without my body, or this sensation without this brain state, or pain without C-fiber firing, are just like this? On this view, the identities posited by the identity theorist are genuine a posteriori necessities, and our intuitions that they are possible false are illusions, to be explained in terms of a confusion of qualitatively identical scenarios with genuine counterinstances.

Kripke thinks not:

> "Now I do not think it likely that the identity theorist will succeed in such an endeavor. ...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 cases a qualtatively analogous statement could be false. . . .
>
> 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 could 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. . . . 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. . . .
>
> The trouble is that the notion of an epistemic situation qualitatively identical to one in which the observer had a sensation S simply it 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." (150-153)

Standard explanations of the illusion of contingency rely on there being a qualitatively identical situation in which one of the key terms — e.g., 'Hesperus', 'gold', 'heat', 'water' — does not refer to what it actually does. But consider a qualitatively identical situation to one in which you are in pain. Could that situation be such that you are not in pain? No. And this is what makes the case of mind-body identities different in principle, in Kripke's view, from other scientific identifications.