Jason Bridges
Phil 340, Winter 2002

Some help with Kripke's modal argument
against the description theory of names

I. The core idea of the description theory of names

We can view the 'description theory of names' as a theory of reference for proper names. A *theory of reference for a type of expression E* is a theory of how in general the references of E's get determined. It tells a story about what makes it the case that particular E's spoken by particular people on particular occasions refer to what they do.

For example, consider *demonstrative expressions*, like "that actor", "this one", or "these children". It seems evident that a plausible theory of reference for demonstrative expressions will say something about the role generally played by the *perceptual salience* of a particular object in determining the reference of an utterance of such an expression. Surely the fact that my utterance of "that actor" refers to a certain man on stage is crucially connected to the fact that I am, or was recently, looking at (or listening to) that man.[1]

We've discussed how the reference of a demonstrative expression like "that actor" is context-dependent: it varies from utterance to utterance depending on context. In contrast, the reference of a proper name is not context-dependent.[2] Hence it is not to be expected that a theory of reference for proper names will appeal to contextual considerations such as which objects are perceptually salient at the time of a given utterance of a name.

So what does matter for the determination of the reference of a proper name? The description theory of names, due to Russell, has at its core the following simple idea. For every name in a speaker S's repertoire, there is some property that S associates with that name, a property that S believes is in fact possessed by one and only one object. The name on S's lips then refers to whatever unique object has that property. If S is mistaken and there isn't one and only one object that has the property—either because nothing does or more than one thing does—then S does not refer to anything.

For example, I might associate the name "Kripke" with the property of being an inventor of possible-worlds semantics, believing that one and only one object has that property. I am, let us suppose, correct in that belief. Hence "Kripke" on my lips refers to whoever it is that invented possible-worlds semantics. On the other hand, suppose I associate the name "Kripke" only with the property of being a philosopher who taught at Princeton, inexplicably believing that only one such philosopher has ever existed. Then "Kripke" on my lips refers to no one.

This idea seems to have a number of things going for it. For one, it appealingly suggests that successful use of a name is a matter of having a piece of knowledge (namely, knowledge that a certain property is possessed by one and only one object). Thus it portrays the use of names as fundamentally a cognitive or intellectual enterprise (more on why this is attractive later). Second, the account provides an explanation of how we can have names for objects that are not found in our environment—indeed, that existed at times and places very distant from our own. It doesn't matter that I've never encountered Homer (or encountered anyone who's encountered Homer); I can refer to him by a name simply because I correctly believe that a unique person wrote the *Iliad*.

---

[1] A general theory of reference for demonstratives, even if it takes this kind of case as basic, will have to acknowledge a range of other possibilities. Sometimes it is not what is currently perceptually salient, but what is remembered, that matters for the reference of an utterance of a demonstrative expression: e.g., "That steak was delicious", said to oneself a week after the fact while in a pleasant reverie. And sometimes we demonstratively refer to objects that are not in our perceptual environment by dint of demonstrating representations of such items, e.g., drawings, photos, and images on TV.

[2] Since more than one person may have the same proper name, proper names, like many common words, can be *ambiguous*. Ambiguity is different from context-dependence. Consider how each might be treated within a Davidson-style theory of meaning.

To recap, here is the core idea of the description theory of names:

> Let N be a name used by speaker S. Then there is some property p that S associates with N, where this involves:
> 1. S believes that one and only one object has p.
> 2. If there is a unique object that has p, N (on S's lips) refers to that object.
> 3. If there is not a unique object that has p, N (on S's lips) fails to refer to anything.

II. The description theory as a theory about the meaning of a proper name

Can we say anything more about what is involved in S's 'associating' N with p other than 1-3 above? The answer is yes. But Kripke shows there are in fact two different ways of expanding the core idea.

Russell himself, and as a consequence nearly all of his followers, conceived of the description theory as a theory about the *meanings* of names. For Russell, the sense in which a speaker associates "Kripke" with the property of being the inventor of possible-worlds semantics is this: for that speaker, "Kripke" and "the inventor of possible-worlds semantics" are *synonyms*. They mean the same thing. Russell even sometimes puts the idea as follows: "Kripke" is short for (or abbreviates) "the inventor of possible-worlds semantics".

What is it for two expressions to be synonyms? One point seems clear: synonyms are intersubstitutable *salva veritate*. That is:

> If e and f are synonyms, then for all possible sentences x containing e, replacing e with f yields a new sentence y that is true iff x is true, and for all sentences x containing f, replacing f with e yields a new sentence y that is true iff x is true.[3]

Thus if "flammable" and "inflammable" are synonyms, we ought to be able to replace "flammable" with "inflammable" in the following sentence, "It's a good idea to give flammable pajamas to a child," and get a sentence that has the same truth-value. And so we can (both sentences are false). Similarly for all sentences involving those expressions.

The point is even more obvious in the case of actual abbreviations. Replacing "North American Free Trade Agreement" with "NAFTA" in a sentence will always preserve truth-value (and vice-versa). Whatever's true of the North American Free Trade Agreement is true of NAFTA.

So conceiving of the description theory of names as a theory about the meanings of names yields a new condition (and motivates us to rewrite the other conditions in terms of descriptions rather than properties):

> The description theory of names *as* a theory of the meanings of proper names:
> Let N be a named used by speaker S. Then there is some description D such that N on S's lips means the same as ('abbreviates') D. That is:
> 1. S believes that one and only one object satisfies the description D.
> 2. If there is a unique object that satisfies D, N (on S's lips) refers to that object.
> 3. If there isn't a unique object that satisfies D, N (on S's lips) fails to refer to anything.
> 4. For all sentences x containing N, replacing N with D in x yields a new sentence y that on S's lips is true iff x is true.

---

[3] This may not hold for certain sentences involving propositional-attitude contexts. We can ignore that complication here.

III. The modal objection

Kripke's modal objection concerns 4. He argues that for almost every name N, there simply is *no* description D such that N and D satisfy 4. Hence, the description theory of names must be false as a theory of the meanings of proper names.

Some Kripkean terminology:
An expression e *designates* an object o iff:
a) e is a proper name and e refers to o; or
b) e is a definite description and o is the unique object that fits the description.
A *designator* is an expression that designates something.

Clearly the only possible name/description pairs that are candidates for satisfying 4 are those in which the name and definite description designate the same object. If N refers to Bob Hope and D is uniquely satisfied by Bing Crosby then, in spite of the myriad similarities between these objects, there will be at least some sentences containing N that change their truth-value when N is replaced with D. (That N and D must designate the same object is also ensured by conditions 1 and 2.)

So let us consider a pair of this sort: "Nixon" and "the 37th President of the U.S.". It can easily seem at first sight that this pair does satisfy 4: "Nixon wrote a number of books" and "The 37th President of the U.S. wrote a number of books" (both true); "Nixon was honest" and "The 37th President of the U.S. was honest" (both false); "Bush belongs to the same political party as Nixon" and "Bush belongs to the same political party as the 37th President of the U.S." (both true); and so on.

But now consider the following:
If Kennedy had won the election in '68, Nixon would not have been the 37th President of the U.S.
If Kennedy had won the election in '68, Nixon would not have been Nixon.
The first of these sentences is true and the second is false. Hence "Nixon" and "the 37th President of the U.S." do not satisfy 4.

Note: The excellent point was made in class last time that there is a way of interpreting an utterance of "Nixon might not have been Nixon" so that what is meant by the utterance could be true. In particular, the utterance could mean that had Nixon not been elected he might not have been so paranoid, blustering and deceitful—so, one wants to say, Nixony. The point is well taken, but irrelevant for two reasons. First, there is the question of whether we want to say that what is *explicitly said* by the utterance is true, or whether we take the true claim simply to be something that the speaker *implies* (i.e., means, intimates) by the utterance.[4] There are good reasons for thinking the latter, and this makes the point irrelevant to our current topic, for it's only what an utterance of a sentence explicitly says that matters for questions of truth conditions and synonymy and the like. Second, if we are prepared to hold that that the second sentence could literally mean that Nixon might not have been paranoid and deceitful, then we must be prepared to regard "Nixon" as semantically ambiguous, in particular, as able to serve both as a name referring to Nixon and as an adjective meaning 'paranoid and deceitful'. It is this second alleged meaning of Nixon that would make the sentence true on the suggestion under consideration. (Maybe if Kennedy had been elected Nixon wouldn't have ended up paranoid and deceitful; maybe Kennedy would have been the one who was so 'Nixon'.)   But it is simply irrelevant to the question of how to understand *proper names* that the sentence is true when the second "Nixon" is read as an adjective.

---

[4] Compare: If I say in a letter of recommendation for a philosophy job, "This candidate has good handwriting," when in fact the candidate does not have good handwriting, what I literally say is false. But what I *mean* (namely, that the candidate has little to offer) may well be true. Remember Grice.

Maybe we simply chose the wrong definite description. Maybe something other than "The 37th President of the U.S." could pair off with "Nixon" to satisfy condition 4.

Kripke has a general argument that this cannot happen. First, more terminology:
     For our purposes, we may think of a *possible world* simply as a way things could be or might have been. Thus:
     "There is a possible world in which S is f" is equivalent to "S might have been f (or could be f)."
     "In relevant possible worlds in which S is f, T is g" is equivalent to "If S were f, T would be g".

     The *actual world* is the way things in fact are. Since the way things are is obviously a way things could be, the actual world is one of the possible worlds.
     e *designates* o *in possible world w* iff in the context of talk about w, e designates o.
     A *rigid designator* is an expression that designates the same object in every possible world.
     A *non-rigid designator* is a designator that is not rigid.

Kripke argues as follows:
I.   All names are rigid designators.
II.   Almost all definite descriptions are non-rigid designators.
III.   Therefore, condition 4 is rarely, if ever, satisfied. (And so the description theory as a theory of the meaning of names is false.)

What is the ground for premises I and II? Consider the following sentence schema:
     There is a possible world in which _____ is not the object that is _____ in the actual world.

Take any designator. If putting the designator in both spaces yields a sentence that is true, the designator is non-rigid. If the sentence is false, the designator is rigid. Why? Well, if there is no possible world in which e is a numerically different thing than what e is in the actual world, "e" designates the same thing in all possible worlds. Thus "e" is a rigid designator. And if there *is* some possible world in which e is a different object from what e is in the actual world, then "e" doesn't designate the same thing in all possible worlds. Thus, "e" is non-rigid.

Let's try "the 37th President of the U.S.":
     There is a possible world in which the 37th President of the U.S. is not the object that is the 37th President of the U.S. in the actual world.

Clearly, this is true. We've already discussed one such possible world: the one in which Kennedy won the election.

Kripke thinks that a little reflection will show that almost all definite descriptions will make the sentence schema true. The reason, intuitively, is that a definite description designates what it does by laying down a condition. Whatever uniquely satisfies that condition is the object designated by the description. But it is in the nature of the vast majority of conditions that what satisfies them will depend on features of how the world is that might have been different. This is certainly true of the condition of being the 37th President of the U.S.: certainly the world might have been different in ways that would alter which object satisfied that condition. Thus what a definite description designates will vary depending on what possible world we are discussing.[5]

---

[5] There are exceptions: some definite descriptions couldn't be satisfied by any other objects than the ones by which they are actually satisfied. Consider: "the natural number between three and five". But such cases are very special and so cannot help the

Now try "Nixon":

> There is a possible world in which Nixon is not the object that is Nixon in the actual world.

Kripke argues that this is obviously false. For what could such a possible world be like? Is it one in which Nixon didn't do any of things we ordinarily ascribe to Nixon? But then notice that in describing that world I used "Nixon" to refer to Nixon. "Nixon" designates Nixon even in a possible world in which Nixon is very different in all kinds of ways from how Nixon actually was. If this weren't so, we couldn't even coherently *talk about* such a possible world.[6]

Note: Many commentators on Kripke spell out "e designates o in possible world w" in something like the following way: if things were as they are in w, e would designate o. This way of putting the idea (contrast the one above) can encourage the mistaken thought that what an expression designates in a possible world is a function of how people *in* that possible world use the expression. This would entail that *no* expression can be a rigid designator, because, obviously, any expression could have been used differently than how it is in fact used in the actual world.

But the explanation of "e designates o in possible world w" just given is misleading. Consider the following sentence:

> Nixon might have been named "Kennedy", and Kennedy might have been named "Nixon".

The fact that this sentence is true *supports* the claim that names are rigid designators. Even though people *in* the described possible world would use "Nixon" to refer to Kennedy, still I, in talking about that possible world, use "Nixon" to refer to Nixon. And it is the latter fact that is relevant to the question of whether "Nixon" is a rigid designator.

We've just argued for premises I and II. What about the inference from I and II to III? It is clearly valid. Suppose D1 and D2 both designate o in the actual world, but that D1 is a rigid designator and D2 is a non-rigid designator. Then D1 and D2 cannot be intersubstitutable *salva veritate*, which is what condition 4 requires. For let W be a possible world in which D2 designates some object p other than o. (There must be such a world if D2 is non-rigid.). And let "is F" be a predicate that is true of o in W but not of p in W. (There must be such a predicate, since all pairs of objects differ in *some* respect.) Then: "In W, ___ is F" is true when D1 fills the blank, but false when D2 fills the blank.

IV. The description theory as a theory about how the reference of a proper name is fixed

Kripke points out that the description theory can still function as a theory of reference for proper names (in the sense described at the beginning) without having to be construed as a theory of the meaning of proper names.

   This version of the description theory goes as follows:

> Let N be a name used by speaker S. Then there is some definite description D that 'fixes the reference' of N for S. That is:
> 1. S believes that one and only one object satisfies D (in the actual world).
> 2. If there is a unique object that satisfies D (in the actual world), N (on S's lips) refers to that object in all possible worlds.
> 3. If there isn't a unique object that satisfies D (in the actual world), N (on S's lips) fails to refer to anything in any world.

---

description theorist, who is looking for a *general* account of how the reference of a name is determined (and is in any case more interested in names of people and other physical objects than numbers or other abstract objects.)

[6] Kripke writes in this vein, "Not only is it true *of* the man Aristotle that he might not have gone into pedagogy; it is also true that we use the term 'Aristotle' in such a way that, in thinking of a counterfactual situation [i.e. possible world] in which Aristotle didn't go into any of the fields and do any of the achievements we commonly attribute to him, still we would say that was a situation in which *Aristotle* did not do these things" (p. 62)

This approach tries to reconcile the description theory with Kripke's claim that proper names are rigid designators. It does so by abandoning the claim that definite descriptions are synonymous with proper names. The reference-fixing role of definite descriptions flows rather from the fact that a speaker intends that whatever object that D designates in the *actual world* be the object that N designate in *all* possible worlds (even in worlds where that object does not satisfy D).

The idea may seem baroque. But Kripke suggests that at least some names really do function this way. Take, for example, the name, "Jack the Ripper". According to Kripke (p. 79), it is plausible to regard this name as having been introduced to the language by the London police via a description, say, "the man who murdered a string of people in Whitechapel in 1888". That is to say, the police introduced the term with the intention that it refer to whoever in fact uniquely satisfied that description. The crucial point is that we can accept this account of how the reference of "Jack the Ripper" is determined while still regarding "Jack the Ripper" as a rigid designator, so long as we take the name to refer in *all* possible worlds to the man who satisfies the relevant description in *this* possible world. For consider:

> In the possible world in which Jack the Ripper convinced his friend Tom to carry out the murders for him, Tom, not Jack the Ripper, is the man who murdered a string of people in Whitechapel in 1888.

This sentence is true. It shows that "Jack the Ripper" is not synonymous with "the man who murdered a string of people in Whitechapel in 1888". But the description theorist can accommodate the truth of this sentence if she abandons the synonymy claim, and holds instead that "Jack the Ripper" refers in every possible world (including the one in which Tom commits the murders) to the man who committed the murders in the actual world.

The example shows that this version of the description theory is not committed to the claim that names and their reference-fixing descriptions are intersubstitutable *salva veritate*. It thereby bypasses the modal objection entirely.

However, Kripke has a number of other objections against this view. The upshot of these objections is not that the account is incoherent, but just that it applies to at best a handful of names in actual use. "Jack the Ripper" is an anomaly.

V. The cluster-of-descriptions theory
So far, we've taken the description theory to hold that each proper name is associated with a single description. This is the classical formulation of the view found in Russell. But Kripke spends most of the time discussing a more complicated version of the theory, in which it is a whole cluster or family of descriptions that is said to be associated with a given name by a given speaker.

So for example I might associate with "Nixon" all of the following: "President of the U.S.", "paranoid and deceitful", "owner of the dog Checkers", "major player of Watergate scandal", "member of secret extraterrestrial cabal". The reference of Nixon is determined as whatever object uniquely satisfies *most* of these descriptions (with some descriptions perhaps being given greater weight in the 'vote').

Just as before, we may understand the theory as saying either that the cluster of descriptions is synonymous with the proper name or that it merely fixes the reference of the name. On the former interpretation, the cluster theory is subject by Kripke to a precise analogue of the objection we discussed against the single-description theory. On the latter, it is subject to objections that we will discuss. These objections aim to show that, although the view is coherent, few names, if any, actually function in the way it suggests.