Reference and Necessity

(1997)

Saul Kripke remarked, at the beginning of his lectures, *Naming and Necessity*, that he hoped his audience would see some connection between the two topics mentioned in his title. In those lectures Kripke defended some bold theses, some about naming that belong to semantics and the philosophy of language, others about necessity that belong to metaphysics. It is clear that the arguments for the different theses were interrelated, but it remains a matter of debate just what the connections are, both in Kripke's argumentative strategy, and in the issues themselves. Kripke and Hilary Putnam were criticized for attempting to derive metaphysical conclusions—about the essential properties of things—from premises in the philosophy of language—about the nature of reference and the semantics of proper names. One might instead think that the direction of Kripke's arguments go the other way: that conclusions about reference and proper names were derived in part from controversial metaphysical assumptions about possible worlds and essential properties. Either way, there is reason to be puzzled: on the one hand, one might be skeptical (to borrow the metaphor that Nathan Salmon used to express his puzzlement about this) that one could, without sleight of hand, pull a metaphysical rabbit out of a linguistic hat.[1] On the other hand, one might wonder why a proper understanding of the way our language happens to work should require controversial assumptions about the metaphysical nature of the world that our language talks about. My aim in this paper is to try to resolve some of this puzzlement by clarifying the relationship between theses and questions about reference and theses and questions about necessity and possibility. In the background of my discussion will be very general questions concerning how claims about the way we talk about the world relate to claims about what the world must be like, but in the foreground will be more specific questions concerning the relations between the different theses Kripke defends about individuals and their names. My main claim will be that Kripke's contribution was not to connect metaphysical and semantic issues, but to separate them: to provide a context in which questions about essences of things could be posed independently of assumptions about the semantic rules for the expressions used to refer to the things, and questions about how names refer could be addressed

---

[1] See Salmon (1981).

without making assumptions about the nature of the things referred to. I will argue that Kripke's theses about proper names and reference do not presuppose any metaphysical theses that ought to be controversial, though even stating those theses requires a framework that might be thought not to be metaphysically neutral. And I will argue that no metaphysical conclusions are derived from theses about reference and names, although clarification of the nature of reference helps in the rebuttals to arguments against metaphysical theses that Kripke defends.

1. Questions and theses

I will start (in section 1) by contrasting three kinds of questions that Kripke discusses in *Naming and Necessity*—two that belong to semantics and the philosophy of language, and one that belongs to metaphysics—and sketching the answers that Kripke defends, along with contrasting answers that he criticizes. Then (in section 2) I will discuss the apparatus that Kripke uses to clarify his questions—the possible worlds framework—and argue that it should be understood not as a metaphysical theory, but as a methodological framework in which alternative metaphysical and semantic theses can be stated. In sections 3–5 I will look in more detail at the arguments for the different theses, and the way the three different kinds of issues interact.

At this point my aim is just to set the stage by making some simple distinctions between questions, and stating, without much explanation or argument, some alternative answers to the questions. First there are questions of what I will call descriptive semantics. A descriptive semantic theory is a theory that says what the semantics for the language is without saying what it is about the practice of using that language that explains why that semantics is the right semantics. A descriptive semantic theory assigns *semantic values* to the expressions of the language, and explains how the semantic values of the complex expressions are a function of the semantic values of their parts. The term 'semantic value', as I am using it, is a general and neutral term for whatever it is that a semantic theory associates with the expressions of the language it interprets: the things that, according to the theory, are the interpretations of simple expressions, and are the arguments and values of the functions defined by the compositional rules that interpret the complex expressions. If, for example, the semantic theory in question assigns senses or intensions to the names and predicates of a language, and explains the senses of complex expressions as a function of the senses or intensions of their parts, then as I am using the term, the semantic values of that semantic theory will be the senses or intensions. The particular descriptive semantic question we will be concerned with is the question, what kind of thing is the semantic value of a proper name?

Second, there are questions, which I will call questions of foundational semantics, about what the facts are that give expressions their semantic values, or more generally, about what makes it the case that the language spoken by a particular individual or community is a language with a particular descriptive semantics. The specific question of this kind that we will focus on is the question, what is it about the situation, behavior, or mental states of a speaker that makes it the case that a particular proper name, as used by that speaker in a particular linguistic community, has the semantic value that it has?

Third, there are questions about the capacities and potentialities of the things in the domain that forms the subject matter of some language; what, for example, might have been true of the things such as persons and physical objects that are the referents of some particular proper names?

Kripke's answer to the first question—the descriptive semantic question about proper names—is the Millian answer: the semantic value of a name is simply its referent. The contrasting answer that he argued against is that the semantic value of a name is a general concept that mediates between a name and its referent—a concept of the kind that might be expressed by a definite description. According to this contrasting answer, the semantic value of the name—its sense or connotation—determines a referent for the name as a function of the facts: the referent, if there is one, is the unique individual that fits the concept, or perhaps the individual that best fits the concept.

Kripke's answer to the second question—the foundational semantic question—is that a name has the referent that it has in virtue of a causal connection of a particular kind between the use of the name and the referent; the referent is the individual that plays the right role in the causal explanation of the fact that the name is being used in the particular context in question in the way that it is being used. In the case of this question, it is less clear what the contrasting thesis is, since the question is not explicitly addressed by the philosophers Kripke is criticizing. But what seems to be suggested is that the sense of a name is determined by the abilities and dispositions of the speaker to describe or identify a certain individual.

In response to the question about the capacities and potentialities of things that we commonly refer to with names, Kripke defends the thesis that it makes sense to talk about the logical potential of an individual thing independently of how it is referred to, and that this potential is greater in certain ways, and less great in others, than some philosophers have supposed. For example, Shakespeare need not have been a playwright; he need not have written anything at all. He might have died in infancy, and been someone who we never heard of—even someone of whom all trace was lost after the seventeenth century. He might not have been called 'Shakespeare' by us, or by anyone. His plays, or at least plays word for word just like his plays, really could have been written by someone else of the same name. On the other hand, Shakespeare could not possibly have had parents other than the ones that he in fact had. In contrast, others have conceded that

Shakespeare might have lacked any one of the attributes commonly attributed to him, but argued that he could not have lacked them all. Attributes not commonly known to apply to Shakespeare, such as having the particular parents that he in fact had, are all attributes that he might have lacked.

One thing I will argue is that while Kripke defends these theses about the descriptive semantics of names, the way the reference relation is determined, and the capacities and dispositions of human beings and physical objects (and I think he makes a persuasive case for each of them), his most important philosophical accomplishment is in the way he posed and clarified the questions, and not in the particular answers that he gave to them. I will suggest that we might buy Kripke's philosophical insights while rejecting all of the theses—while opting for a pure Russellian description theory of ordinary names, a non-causal account of the way names get associated with their values, and in metaphysics either for an anti-essentialist thesis according to which Shakespeare might have been a lamppost or a fried egg, or for a Leibnizian theory according to which Shakespeare had even his most apparently accidental properties essentially. The positive case for the theses that Kripke defends is not novel philosophical insight and argument, but naive common sense. The philosophical work is done by diagnosing equivocations in the philosophical arguments for theses that conflict with naive common sense, by making the distinctions that remove the obstacles to believing what it seems intuitively most natural to believe.

2. The possible worlds framework

To accept what I will argue is Kripke's main philosophical contribution, you do have to buy a framework, an apparatus that he used to sharpen and clarify the contrasting theses, both semantic and metaphysical. I won't try to claim that the apparatus is either semantically or metaphysically neutral, but I will argue that the motivation and commitments of the framework are more methodological and conceptual than they are metaphysical. Philosophers often talk as if the decision to theorize with the help of this framework is, if one takes the claims one makes while using it seriously, a specific ontological commitment to a certain kind of entity. Some philosophers reject the framework because they reject the ontological commitment that its serious use makes. One hears: 'I don't believe in possible worlds,' as one might hear people say that they don't believe in transubstantiation, or flying saucers from other planets (commitments that some philosophers believe are about as plausible as the commitment to possible worlds). I think this attitude is based on a misconception (although I have to concede that it is a misconception that some defenders of possible worlds share with the critics). It is not that it is a misconception to think that serious talk about possibilities commits one to the existence of the possibilities one claims there are, just as it is not a misconception to think that the literal use of

quantifiers commits one to the existence of things that one purports to quantify over. But it is not the framework itself that makes the specific commitments, just as it is not the semantics for first-order logic that makes any particular ontological commitment. Suppose someone were to reject the standard (extensional) semantics for first-order logic on the ground that he did not believe in individuals, to which that semantics is ontologically committed. The proper response would be to point out that first-order semantics is a framework for doing ontology, and not a particular thesis about what ontology is correct. Individuals are whatever there is to talk about; the semantic theory itself says nothing about what there is to talk about, and so makes no particular ontological commitments. Quine's slogan, 'To be is to be the value of a bound variable,' is not an ontological thesis, but an attempt to promote a framework in which the ontological commitments of alternative philosophical and scientific theories can be stated without equivocation and compared.

The Leibnizian slogan, 'necessity is truth in all possible worlds,' should be understood in a similar spirit. This slogan and the possible worlds framework that it presupposes should, I think, be understood not as an attempt to provide an ontological foundation for a reduction of modal notions, but as an attempt to formulate a theoretical language in which modal discourse can be regimented, its structure revealed, equivocation diagnosed and avoided. Modal discourse—speech that involves words such as 'may', 'might', and 'must'—is notoriously complex and problematic, providing fertile ground for ambiguities, both ambiguities of scope that arise because the semantic structure of modal statements is complicated and not simply reflected in surface syntax, and ambiguities that arise from alternative senses and context dependence of the modal words. Modal words interact with each other and with quantifiers, descriptions, temporal modifiers and grammatical tense, aspect, and mood in complicated ways that are difficult to sort out. Philosophical puzzles about, for example, necessary connection and counterfactual dependence, reference to nonexistent things, capacities and dispositions, the openness of the future, the necessity of the past and the ability to do otherwise, will presumably not all be dissolved simply by getting clear about modal discourse, but everyone should agree that clarifying the discourse in which such problems are posed is an essential first step.

It is important to separate disagreements based on contrasting interpretations of the way the language works from disagreements about the claims that the language is being used to state. Whatever one's metaphysical beliefs about the reality that modal discourse purports to describe, one should agree that it would be nice to have a language that is free of some of the ambiguities that infect modal discourse, and in which the claims made with modal words and constructions might be paraphrased—a language that uses only parts of discourse that are relatively clear and uncontroversial (the indicative mood and quantifiers), but that still has the expressive power to

make claims about what might, would, or must be true. Achieving such clarification does not require a reductive analysis of modal to nonmodal concepts, and so it is not required that a canonical language in which we do modal semantics be built on some pure, nonmodal foundation, any more than formal languages designed to clarify quantification needed to be built on some pure, nonquantificational foundation (whatever that would be). What is needed is only the kind of opportunistic departure from ordinary language involved in the bootstrap operation that Quine called 'regimentation.' In the kind of regimentation Quine recommends, we begin with *ad hoc* paraphrases to remove ambiguity; we introduce variables to facilitate cross-reference, and we adopt a syntax in which quantifier scopes are reflected in a simple and systematic way in the order of the symbols and the placement of parentheses. 'The artificial notation of logic,' Quine remarked, 'is itself explained, of course, in ordinary language.'[2] Similarly, the primitive resources of the possible worlds framework are explained in ordinary modal language, and the explanations will be intelligible only to one who understands at least some of that part of language. A modal skeptic who doubts that anything both meaningful and true is said in modal discourse will doubt both the value and the intelligibility of a framework in which that discourse is clarified. But it may be that the source of the skepticism is in the equivocations and unclarities that the framework helps to remove.

The general strategy is to find a part of our modal discourse that seems relatively free of the particular equivocations and unclarities that infect modal discourse generally, a part that might be developed and used to clarify the rest. We look for a way of making modal claims that uses paraphrases to avoid problematic constructions, a way that uses forms of expression that may perhaps be stilted and less idiomatic than the familiar ones, but that will still be recognizable paraphrases of ordinary modal claims. The following assumption about what is, in any sense, possible points the way to one such strategy of paraphrase: if something might be true, then it might be true in some particular way. It would make no sense to affirm, for example, that there might be life elsewhere in our galaxy, while denying that it is possible that there be life in any particular part of the galaxy, or that there might be life of any particular kind—animal life, or nonanimal life—elsewhere in the galaxy. If something is possible, then it is possible that it be realized in a concrete way—perhaps in many alternative ways. The possible worlds framework begins with this simple assumption, and with the assumption that, in general, statements about what may or might be true can be described in terms of the ways a possibility might be realized. The framework takes alternative specific ways that possibilities might be realized as the primitive elements out of which propositions—the things that are said to be possible, necessary, or true—are built and in terms of which the

2 Quine (1960: 159).

modal properties of those propositions are defined. The main benefit of this move is that it permits one to paraphrase modal claims in an extensional language that has quantifiers, but no modal auxiliaries, and so in a language in which the semantic structure of the usual modal discourse can be discussed without begging controversial questions about that structure.

I have been arguing for the metaphysical neutrality of the possible worlds framework, but I should emphasize that I do not mean to suggest that the use of the framework is free of ontological commitment to possibilities (ways things might be, counterfactual situations, possible states of the worlds). Regimentation clarifies one's commitments, but does not pretend to eliminate them. Furthermore, it must be conceded that the moves made in this regimentation of modal discourse (particularly the move that paraphrases '___ might have been true in various particular ways' as 'there are various particular ways that ___ might have been true') are not completely innocent. As Quine would be the first to emphasize, no strategy of regimentation is neutral in any absolute sense: 'The quest of a simplest, clearest overall pattern of canonical notation is not to be distinguished from a quest of ultimate categories, a limning of the most general traits of reality.'[3] But it is a desideratum of any such project that it be able to accommodate and articulate a range of alternative responses to the questions and puzzles that motivated the project. I think the possible worlds framework satisfies this desideratum, but the real test of this claim is not in some general methodological argument, but in the fruits of the work that is done with its help.

3. What are the semantic values of names?

The possible worlds framework provides the resources to state and clarify both metaphysical and semantic theses. In both cases, I want to argue, the principal conceptual benefit of the apparatus is that it provides an account of a subject matter that is independent of languages used to describe that subject matter. Of course whether the subject is geology or modal metaphysics, we never get away from language—it is just too hard to say very much without using it. But just as we want to distinguish rocks from words (even if we have to use words for rocks to do it), so it is useful to distinguish possibilities from the words used to describe them. To make this distinction is not to beg any questions against the philosophical thesis that the source of all necessity is in language; a conceptual distinction does not foreclose the possibility that one thing distinguished may in the end be reduced to the other.

To see that possibilities are part of the subject matter of semantics as well as of modal metaphysics one need only make the following assumptions: first, a central function of an assertion is to convey information, and

3 Ibid. 161.

information is conveyed by distinguishing between possibilities. Second, a principal goal of semantics is to explain how the expressions used to perform speech acts such as assertion are used to convey information—to distinguish between possibilities—and how the way complex sentences distinguish between possibilities is a function of the semantic values of their parts. To understand what is said, for example, in an utterance of 'The first dog born at sea was a basset hound' one needs to know what the world would have to be like in order for what was said in that utterance to be true.

These simple assumptions about the goal of semantics might be expressed in terms of truth-conditions: semantics is concerned, among other things, with the truth-conditions of statements, and the way their truth-conditions are a function of the semantic values of their parts (where semantic values are whatever they must be in order to contribute appropriately to truth-conditions). What are truth-conditions? If we are looking for an answer to this question that identifies a nonlinguistic object that semantics can associate with statements, it seems natural to say that the truth-conditions of a statement are the possibilities that, if realized, would make the statement true. We want a conceptual distinction between truth-conditions and any particular forms of expression in which those truth-conditions might be stated simply because it is useful to theorize about a language in a different language, and when we do so, we want to be able to talk, not just about the language we are theorizing about and the world, but about the relation between the language we are theorizing in, but about the relation between the language and the world.

The task of descriptive semantics, in this framework, is to say what kinds of things the semantic values of expressions of various categories are, and to explain how the truth-conditions of sentences (or sentences in context) are a function of the semantic values of their constituents. So to give the semantic value of a proper name is to say what contribution a proper name makes to the truth-conditions of the sentences containing it, where the truth-conditions of a sentence, or a sentence in context, are represented by the set of possibilities that, if realized, would make what the sentence says in that context true. The two answers that Kripke compares—the Millian and the Fregean answers—are made precise in the following ways. (1) The semantic value of a name is simply its referent; the proposition expressed by a simple sentence containing a name is the proposition that is true in a possible world if and only if that referent has the attribute expressed by the predicate of the sentence. (2) The semantic value of a name is its sense, which is a concept that applies to at most one individual in each possible world (the kind of concept that might be expressed by a simple sentence containing a name is the proposition that is true in a possible world if and only if the individual to which the concept applies in that world has the attribute expressed by the predicate of the sentence.

4. How do names get their semantic values?

Why do Searle and Dummett think that we could not speak a language of the kind that Kripke described in which the semantic value of a name is

Thus far, I have talked only about a question and a framework for clarifying alternative answers to it, and not about arguments in support of one or the other of the answers. The framework is neutral on the question of which of these alternatives, if either, gives a correct account of the semantics of these expressions in English and other natural languages that we identify as proper names. This seems to be an empirical question of no particular philosophical interest, a question that philosophical analysis and argument are not relevant to answering. The way in which the alternative answers are articulated in the framework does point the way to some of the empirical considerations that may be relevant to settling the issue by making clear what the consequences of those alternatives are, but it appears that no philosophical—certainly no metaphysical—issue hangs on which answer is right. Even though Kripke defended, on empirical grounds, the Millian answer, he nowhere suggested that things had to be this way. For all he argues, we might perfectly well have spoken a language with names that all referred only by having senses that determined referents. It just happens that we do not.

What needed philosophical defense was not the empirical adequacy of the Millian answer, but its coherence. While Kripke did not suggest that philosophical argument could establish that the answer he favored was correct, he had to answer philosophical arguments that purported to establish that it was incorrect. John Searle, for example, argued that

the view that there could be a class of logically proper names, i.e. expressions whose very meaning is the object to which they are used to refer, is false. It isn't that there just do not happen to be any such expressions: there could not be any such expressions.' The view that proper names are 'unmeaning marks,' that they have 'denotation' but not 'connotation,' must be at a fundamental level wrong.[4]

Michael Dummett makes a similar claim: 'there cannot be a proper name whose whole sense consists in its having a certain object as referent, without the sense determining that object as referent in some particular way.'[5] These claims are puzzling, in the light of Kripke's way of posing the problem of the descriptive semantics of proper names. It appears that he showed, simply in setting up the alternatives, how to give a coherent specification of a language containing 'expressions whose very meaning is the object to which they are used to refer.' What kind of argument could show not only that we do not in fact speak such a language, but that we could not possibly do so? To address this question, we need to turn to the second of the two kinds of issues in the philosophy of language that Kripke is concerned with.

[4] Searle (1969: 93).   [5] Dummett (1973: 232).

simply its referent? Searle's reason was that 'if the utterance of the expressions communicated no descriptive content, then there could be no way of establishing a connection between the expression and the object; no way to answer the question, 'what makes *this* expression refer to *that* object?'[6] Dummett's reason is similar: 'an object cannot be recognized as the referent of a proper name . . . unless it has first been singled out in some definite way'.[7] In both cases, the reason for rejecting the possibility of a certain descriptive semantic thesis appeals to considerations that relate to the foundational question, which asks what it is about the capacities, customs, practices, or mental states of a speaker or community of speakers that makes it the case that an expression has the semantic value that it has. What seems to be suggested is that the hypothesis that a language has a Millian semantics poses a foundational question that cannot be given a satisfactory answer.

But this is not the way either Searle or Dummett put their claims, since they do not separate the two questions, 'what is the semantics for names (or the semantic value of a particular name) in the language we speak?' and 'what makes it the case that the language we speak (or a particular name in that language) has this semantics?' Once the two questions are separated, it is difficult to see what could rule out the *possibility* that we speak any language that has a well-defined semantics. If a Millian semantics for names can be articulated, why can't a community of speakers adopt the convention to speak such a language?

The assumption implicit in the rejection of the possibility of a Millian semantics is that the two questions we have separated should receive a single answer. Something like a Fregean sense should explain why a name has the particular referent it has, where this is interpreted to mean that it should explain both what it is about the capacities and attitudes of the speaker that give the name the referent it has, and also what it is that the speaker communicates or conveys in using the name. Kripke charged Frege with conflating these two questions: 'Frege should be criticized for using the term "sense" in two senses. For he takes the sense of a designator to be its meaning and he also takes it to be the way its reference is determined. Identifying the two, he supposes that both are given by definite descriptions.'[8] Whether Frege is responsible for making this mistake is a question of textual interpretation that I will not comment on, but it is clear, I think, that Searle is concerned with both kinds of questions, and that he takes himself to be following Frege. Searle describes his axiom of identification, a principle that is supposed to be constitutive of singular definite reference, as 'a generalization of Frege's dictum that every referring expression must have a sense'.[9] The principle is about what must be communicated or conveyed

6 Searle (1969: 93).
8 Kripke (1972: 59).
7 Dummett (1973: 232).
9 Searle (1969: 80).

(or at least 'appealed to' or 'invoked') in the utterance of the referring expression, but it is also an attempt to say what it is about the capacities of speakers that explains why their referring expressions have the referents they have. 'What I am trying to get at,' he says, 'is how noises identify objects.'[10]

If we are implicitly looking for a semantic account of names that answers both questions at once, then the Millian theory that says that the semantic value of a name is simply its referent looks like a nonanswer; it seems to be denying the obvious fact that there must be something about the capacities, behavior, or mental state of the users of the name that make it the case that the name has the referent that it has. On the other hand, the conflation of the two questions masks the fact that the sense theory, interpreted as an answer to the question of descriptive semantics, is also a nonanswer to the foundational question. Suppose we were to accept the Fregean thesis that names have the referent that they have because they have a sense that determines a function whose value (at the actual world) is that referent. This simply raises the question, what is it about the capacities, behavior, or mental state of the users of the name that makes it the case that the name has the sense that it in fact has? Whether one is a Fregean or not, the two questions need to be distinguished, and once they are, the way is opened for answers to each that are less easily seen as possible answers to the other: the Millian answer to the descriptive question and the causal account of reference that Kripke defends as an answer to the foundational question. This latter thesis—that what makes it the case that a name has a certain individual as its referent is that the individual plays a certain role in the causal or historical explanation of the speaker's use of the name—makes no sense as an attempt to specify a semantic value, a candidate to be the meaning or sense of a name, and so it can be taken seriously only after the two questions are distinguished.

Kripke and other defenders of a causal theory of reference were criticized for the vagueness of their thesis. Causal connections are ubiquitous, and it is obvious that there are a great many individuals that are causally implicated in the speaker's use of the name, but that are not by any stretch of the imagination plausible candidates to be the referent. A proper causal theory of reference would have to specify just what sort of causal connection is necessary and sufficient for reference, and that is a notoriously difficult demand. Kripke himself emphasized that he was presenting not a reductive analysis of reference, but only an alternative picture. To some skeptics, this sounded like an evasion. The suspicion was that any sufficiently specific and precise version of the causal theory would be subject to as many counterexamples as the description theory, and so that the plausibility of the positive alternative to the description theory rested on its lack of specificity. But I think this line of criticism misses the point. What is essential to the

10 Ibid. 83 n.

alternative picture was the separation of the questions, and the distinction between two different ways in which the extension of an expression might depend on the facts: first, what semantic value an expression has depends on the facts; second, if the semantic value is a sense, the extension of an expression with a given semantic value may depend on the facts. The philosophical work was done in making the distinctions that removed the obstacles to accepting the naive answers to the questions that were distinguished. If we ask 'what does one have to know to understand a name?', the naive answer is that one must know who or what it names—nothing more. (In contrast, no one would be tempted to give this answer to the analogous question, what must one know to understand a definite description.) And if we ask, how does a name come to name what it names—what, for example, makes 'Shakespeare' as we use it a name of the particular person Shakespeare?—I think most people would be inclined to point to an historical narrative: his family was called 'Shakespeare,' or something like that, at the time, and knowledge of him, his plays, and his name were passed down through the generations to us. This is not a particularly exciting philosophical theory, but it doesn't seem wrong, and it does seem incompatible with the kind of answer implied by a description theory. Kripke's causal chain story is just an articulation of the naive answer, one that does not add a lot of constructive detail to it. But by separating and clarifying the questions that these naive answers are answers to, he brought out why the theoretical reasons for resisting those answers are bad reasons.

The diagnosis of equivocation is rarely the end of the matter in a philosophical argument. I will sketch a line of argument for the impossibility of a Millian semantics that recognizes the distinction between the two kinds of questions. I don't think this line of argument is successful, but it is instructive to see where it takes the debate.

If there is a credible defense of the thesis that a Millian language is impossible, I think it must challenge the assumption that any well-defined semantics might be the semantics of a language that is used by a speaker, or realized in a speech community. Here is one way that the assumption might be challenged. First, the following seems a reasonable general constraint on the correctness of a claim that a certain semantics is the semantics for the language spoken in a certain community of speakers: if the semantics is correct, then speakers must know, at least for the most part, what, according to the semantics, they are saying. A notion of saying might allow that in some cases one succeeds in saying things using words one does not understand, but it is hard to deny, first that if one doesn't know what one is saying, then one does not mean what one says, and second, that according to a correct account of what people say in a given speech community, speakers generally mean what they say (not in the sense that what they say is sincere—believe what they say—but just in the sense that what they say coincides with what they mean).

Second, we may note that it is possible to give a determinate specification of a semantic value without knowing what that value is, even without anyone knowing what the value is. Consider this example discussed by Gareth Evans: let 'Julius' be a (rigid) proper name for the person (whoever he or she was) who invented the zip.[11] Then (assuming that some particular single individual invented the zip) a sentence such as 'Julius was born in Minsk' expresses a determinate proposition about a particular individual, but we don't know who the individual is, so we don't know what proposition it is that is expressed. We understand a description of the proposition, and we understand, and may believe, metalinguistic statements about that proposition, such as 'What is said by "Julius was born in Minsk" is probably false.' We can have beliefs and make assertions about the truth or falsity of whatever proposition is expressed, but (according to this line of argument) we do not thereby assert or believe that Julius was or was not born in Minsk, and we cannot do so unless we know who invented the zip.

Now suppose that one could make a case that our mental relations to particular individuals are all in relevant respects like our relation to Julius—that since we can know individuals only by description—only as whoever or whatever it is that is presented to us in a certain way—we don't ever know, in the relevant sense, who or what it is that we are referring to with the names we use. Then it would seem to follow that, although we can define a language with a Millian semantics, we could never speak one, since we could not have the knowledge required to know what the sentences of such a language say.

I think the first premise of this argument should be conceded: a semantics for the language spoken by a community of speakers cannot be right if it implies that speakers generally do not know what they are saying. It should also be conceded that according to the Millian semantics for names, as contrasted with the Fregean semantics, speakers do not know what they are saying when they use a name if they do not know who the referent of their name is. But what is it to know what one is referring to? At this point, the battleground shifts from semantics and the philosophy of language to the philosophy of mind where variations on some of the same battles are fought.

Underlying the contrasting answers to the foundational question about reference (what makes it the case that a name has the referent it has?) are contrasting strategies for answering parallel foundational questions about mental states: what makes it the case that a thought—a judgment or an intention—has the content it has, or is about what it is about? The argument sketched above against the possibility of a Millian semantics for an actual language rests on the assumption that thoughts can be about particular things only by expressing general concepts that apply to those particular things.

11 Evans (1982: 31).

individuals. If this were right, then the kind of causal chain or historical explanation story that Kripke and Keith Donnellan told to answer the foundational question about semantics would be an answer that detached the determination of the semantic values of expressions from the mental states and capacities of the users of the expressions, and so would be an answer that was vulnerable to this argument. But why should we think that thoughts, any more than names, can be about individuals only by expressing general concepts?

'What is it,' Searle asks rhetorically in defense of his principle of identification and his argument against the Millian theory,

to *mean* or *intend* a particular object to the exclusion of all others? Some facts incline us to think that it is a movement of the soul—but can I intend just one particular object independent of any description or other form of identification I could make of it? And if so, what makes my intention an intention directed at just *that* object and not at some other? Clearly the notion of what it is to intend to refer to a particular object forces us back on the notion of identification by description.[12]

The suggestion implicit in these rhetorical questions is that an intention to refer to a particular individual must be explained as a behavioral capacity, the capacity to give a general description, or otherwise identify an individual who is, by fitting the description or being the object identified, the object meant or intended. The only alternative, it is implied, is some kind of obscurantist intentional magic, some kind of movement of the soul.

Even if Searle were right in his suggestion that intentions and other intentional states directed at particular individuals must be explained in terms of the capacities of the agent to identify the individual, this would still not give him the additional premiss needed for the argument against the possibility of a Millian semantics for names. For Searle is not arguing that we cannot have intentions and other attitudes toward particular individuals; he is only arguing for a condition that is necessary for having such intentions and attitudes. What he offers instead is a constraint on the content of the attitudes we can have, but what he needs for the argument is a constraint on the conditions under which one can have attitudes with a certain kind of content. Whatever it is that constitutes intending and having knowledge and beliefs about a particular object, 'to the exclusion of all others', so long as it is possible to have such intentions, knowledge, and beliefs, it will be possible to understand and speak a language with a Millian semantics.

But in any case, there is no real argument for the conclusion that mental magic is the only alternative to an explanation of intentionality in terms of an agent's capacities to identify. A causal account of intentional content—an explanation that looks back to how mental states came to be rather than

---

[12] Searle (1969: 87).

only forward to what those states dispose the agent to do—is equally compatible with a nonobscurantist account of intentionality. A causal account of intentions and beliefs seems in fact to be presupposed by the defense of a causal account of reference given by Kripke and Donnellan, since it is argued that speakers not only can refer, but can intend to refer, to particular individuals without being able to describe or identify those individuals. Causal and noncausal accounts of how names get their reference can share the assumption that reference is determined by intentions. The causal theory of reference is causal because it assumes a causal account of the content of the intentions that determine reference.

An argument of Michael Dummett's for the impossibility of a Millian semantics, like the argument I have sketched, bases this conclusion on the impossibility of a certain kind of knowledge: what Dummett calls *bare knowledge of reference*. Here is his characterization: 'A *bare knowledge of the reference of the name a* will consist . . . in knowing, of some object, that *a* refers to it, where this is a *complete* characterization of this particular piece of knowledge.'[13] I am not sure what Dummett means by a 'particular piece of knowledge,' or what it is for a characterization of such a piece to be complete, but if we interpret Dummett's arguments in the context of the possible worlds conception of content, I think it is reasonable to identify his notion of bare knowledge of reference with knowledge of a singular proposition—the proposition that is true in a possible world if and only if a certain particular individual is the referent of the name *a*. An essential step in Dummett's argument for the impossibility of bare knowledge of reference is a claim that is essentially equivalent to Searle's principle of identification: we cannot have what Dummett calls knowledge-what—knowledge of a certain individual that it has some property *F* (for example, knowledge of a certain individual that it is the referent of the name *a*)—unless we have the capacity to describe or identify the object. More strongly, for any true knowledge-what ascription, there must be a true propositional knowledge ascription whose content is a nonsingular proposition that makes the method of identification explicit, and that entails the knowledge-what ascription: a propositional knowledge ascription on which the knowledge-what ascription 'rests.'[14] Now I am not persuaded that this claim is correct, but even if this much is granted, I don't think it gives one reason to reject the possibility of knowledge of singular propositions. Suppose we grant that one cannot know of some particular individual *x* that it is *F* unless for some *G* one identifies *x* as the *G*, and knows that the *G* is *F*. Further, suppose we grant that in a particular case the claim that *y* knows of *x* that it is *F* rests on, and is entailed by, the claim that *y* knows that the *G* is *F*. What has been granted is a claim that certain conditions are necessary, and others sufficient, for having knowledge of a certain kind; but nothing follows from this

---

[13] Dummett (1991: 127).   [14] Ibid. 130.

about the content of that kind of knowledge. If 'bare knowledge' of some object that a refers to it is taken to mean knowledge that can exist in isolation—without knowing anything else about what a refers to, then we can grant that bare knowledge of reference is impossible, but that does not imply that knowledge of x that a refers to it is not knowledge of a particular proposition—a singular proposition—or that it is not possible to have knowledge of such propositions.

We can agree with Dummett that it is a difficult problem to say just what conditions must be met for one to know who or what the referent of a name a is, and that the problem is not solved simply by saying that to have such knowledge-what is to know a singular proposition of the form 'x is the referent of a'. But the problem is not that saying this would be wrong; it is just that specifying the content of a knowledge ascription is not the same as saying what it is for that knowledge ascription to be true.

The distinctions, both on the level of speech and on the level of thought, between questions about what content is and questions about how content comes to be determined, help to open up a place in conceptual space for a causal account of reference and of intentionality generally, and provide a rebuttal to arguments for the impossibility of a Millian semantics for a realized language. They do not, of course, end the debate. Once theoretical obstacles to such accounts are removed, examples and untheoretical considerations make a strong prima facie case for the claim that some such account is correct, but as Gareth Evans reminded Kripke, 'the deliverances of untutored linguistic intuition may have to be corrected in the light of considerations of theory.'15 I think more theoretical considerations also support causal accounts of intentionality, but that is a different part of the story. I want to turn back now to questions about the relation between reference and metaphysical necessity.

5. Names and essences

Whatever the fate of the debate between them, we have a stark contrast between two pictures of the way we are related, both by speech and by thought, to particular things in the world. On one picture, we can think and talk directly about particular things in virtue of our causal interaction with those things; on the other, our mental and linguistic acts relate us to particular things only by our grasping and expressing purely qualitative concepts that may be instantiated by particular things. The question now is: do these pictures of mental and linguistic representation either presuppose or support some particular conception of the nature of the particular things that we talk and think about, or that instantiate our concepts? Specifically, does the conception of reference that Kripke argues for presuppose or

15 Evans (1982: 76).

support the particular brand of essentialism that he defends? The two kinds of issues, I will argue, are independent. The only role of the theory of reference in Kripke's arguments for metaphysical conclusions is to help diagnose and rebut fallacious arguments that rest on a conflation of the two kinds of issues.

One of Searle's arguments against the Millian account of proper names was that it (or at least an 'uncritical acceptance' of it) leads us into some 'metaphysical traps.'16 It is suggested that this conception of proper names presupposes 'a basic metaphysical distinction between objects and properties or aspects of objects'.17 Actually Searle's attitude toward the relation between the metaphysical and the semantic issues is not entirely clear. Does he locate the mistake in the premiss—the Millian semantics for names—or in the inference from this premiss to a metaphysical conclusion? On the one hand, we are warned against 'the original sin of metaphysics, the attempt to read real or alleged features of language into the world'; and 'the metaphysical mistake of deriving ontological conclusions from linguistic theses',18 but on the other hand, the fact that the Millian account of names seems to presuppose this metaphysical distinction is part of an argument against that semantic account. It cannot be a good argument against a semantic account of proper names that someone has illegitimately drawn metaphysical conclusions from that account. So perhaps the view is that the Millian theory of names is already a covertly metaphysical thesis—that false metaphysical conclusions are validly drawn from it because the thesis is the result, rather than the occasion, for reading alleged features of language into the world. But if so, can't we separate the semantic from the metaphysical aspects of the thesis, and evaluate the semantic thesis independently of the metaphysical conclusions that are illegitimately drawn from it? If we couldn't make such a separation, then it would not be so clear that it was illegitimate to draw ontological conclusions from linguistic theses.

Searle's target is the Wittgenstein of the Tractatus, and not Kripke (whose lectures were given after Searle's book was published). But Kripke does make the kind of metaphysical distinction between objects and properties that Searle rejects, and of course he also defends the account of names that Searle argues is the illegitimate source of the metaphysical distinction. I will sketch the way Kripke makes the metaphysical distinction and then argue that his metaphysical theses are compatible with the Fregean picture of mental and linguistic representation, and so do not presuppose the Millian semantics or the causal theory of reference. What is presupposed in the defense of Kripke's metaphysical conception is only that the two accounts of reference and intentionality not be conflated. I will conclude by considering whether there is a dependence that goes in the other direction: whether the semantic picture that Kripke defends presupposes his metaphysical theses

16 Searle (1969: 163).   17 Ibid. 164.   18 Ibid.

about the relation between individuals and their properties. Here the issues are harder to disentangle, but I will argue that the Millian theory of names and the causal theory of reference are compatible with alternative metaphysical conceptions of individuals and their properties. There is no derivation of metaphysical conclusions from semantic premises.

What is it to make a basic distinction between objects and properties? Searle derided the metaphysical picture of an object as 'a combination of its propertyless self and its properties', as well as the contrasting picture of an object as nothing but 'a heap or collection of properties',[19] but what do these pictures really come to? Kripke also scorned the same two contrasting pictures, rejecting the assumption that objects are some kind of 'bare particulars' or 'propertyless substrate underlying the qualities', as well as the claim that they are nothing but bundles of qualities, 'whatever that may mean'.[20] The possible worlds framework suggests a way to express the idea that a particular is conceptually separable from its properties without relying on the rejected picture of a bare particular. Properly understood, the issue concerns the modal properties of an individual. Intuitively, it seems clear that ordinary things might have had different properties from the ones they in fact have: Shakespeare might not have written plays, which, in the possible worlds paraphrase, is to say that there is a possible world in which Shakespeare did not write plays. This possible world is one in which the very person who actually wrote the plays we know and love exists, and did not write plays. So (assuming the modal claim is right) the property of writing plays is not essential to a particular person who has this property. The same goes for lots of other ordinary properties ascribed to ordinary persons and things, but not to all of them. Shakespeare obviously could not have been someone other than Shakespeare (although he could have been called something else), and according to Kripke he could not have been a member of a different species, or even have had different parents. Now these simple modal claims say nothing about names or reference, but in stating them I am using the proper name 'Shakespeare', so the content of what I am saying about counterfactual possibilities might be thought to depend on the semantics for names. If 'Shakespeare' were an abbreviation for a definite description as Russell argued, then the statement that Shakespeare might not have written plays, and its paraphrase, that there is a possible world in which Shakespeare did not write plays, would both be ambiguous. If, for example, 'Shakespeare' were an abbreviation for 'the most famous Elizabethan playwright', then on one interpretation the claim that Shakespeare might not have written plays would be the claim that there is a possible world in which the person who is the most famous Elizabethan playwright in that world did not write plays. This, of course, is not the claim that Kripke intended to make, and it does not seem, intuitively, that the words used to

---

19 Searle (1969: 163).

20 Kripke (1972: 52).

make the claim are open to such an interpretation. The modal claims, in either their ordinary form or in their possible worlds paraphrases, do not seem ambiguous, but that is a linguistic intuition which is separable from the modal intuition about the person who is the referent, by whatever means, of the name 'Shakespeare', the proposition that is expressed by the other reading of the statement that the most famous Elizabethan playwright might not have written plays. Now on what I have been calling the Fregean conception of mental representation, perhaps we cannot have such modal intuitions about particular individuals since perhaps on that conception the only way we can have any thoughts at all about an individual is to have beliefs about whoever it is that is presented to us in some particular way. But this does not matter to the issue, since the metaphysical intuition can be expressed in a perfectly general way: whoever the person is who fits our Shakespeare concept, that person might have been someone who didn't write plays. So the modal theses stand or fall independently of the success or failure of the defense of theses either about the semantics for proper names, or about the way our thoughts relate us to the individuals that are the referents of those names.

If both the Fregean conception of reference and intentionality that Searle favors and the alternative semantic conception that he criticizes are compatible with the metaphysical distinction between particulars and their properties, what is the source of his objections to this distinction? I think they derive from a conflation of the two semantic conceptions. By equivocating between the two semantic theses, one can argue from semantic assumptions to a metaphysical conclusion: On the Fregean picture, there is an analytic, and so necessary, connection between the name 'Aristotle' and a cluster of properties, and so it is legitimate to conclude, as Searle does, that 'it is a necessary truth that Aristotle has the logical sum [inclusive disjunction] of the properties commonly attributed to him.'[21] This claim by itself implies nothing about what must be true of the person Aristotle and so raises no problems about the traditional distinction between objects and their properties. On the unequivocal Fregean picture, 'Aristotle' means, roughly, whoever satisfies the cluster; so according to this semantic hypothesis, if the person Aristotle hadn't satisfied the cluster, he would not have been Aristotle, but he might still have existed. It is only when the Fregean premiss with the incompatible assumption that the name 'Aristotle' is a Millian name that one can take the next step in the argument, the inference from the claim, *necessarily Aristotle satisfies the cluster*, to the conclusion that it is *necessarily true of Aristotle* that he satisfies the cluster. ('I wish to argue that though no single one of them is analytically true of Aristotle, their disjunction is.'[22]) It is only when one has the conclusion that the person Aristotle is necessarily connected with the properties used to

---

21 Searle (1969: 173).

22 Ibid. 169.

identify him that one has reason for skepticism about substance, and for the claim that 'it is misleading, if not downright false, to construe the facts which one must possess in order to refer as always facts *about* the object referred to, for that suggests that they are facts about some *independently* identified object.'[23]

All Kripke does to make a positive case for his modal theses about individuals is to develop the framework in which the theses can be formulated clearly and separated from theses about names and reference. The rest of the work is done simply by pointing out what seems from an intuitive point of view obviously true, once it is clear what the alternatives are. If Kripke's rhetorical style had been a little different, he might have made this point by saying that he was just assembling reminders, not putting forward theses. What philosophy does, he might have said, is simply to put everything before us.[24]

Not all of the metaphysical claims about individuals that Kripke defends on intuitive grounds are equally compelling. On the one hand, it seems hard to deny that we can make intuitive sense of questions about the potentialities of particular individuals, independently of the means used to refer to them. To suppose that Shakespeare never wrote any plays is to envision a counterfactual situation in which Shakespeare—the man himself—wrote no plays. Other theses Kripke defends are more controversial from an intuitive point of view, particularly theses that deny that certain things are possible, or equivalently, that affirm that particular things have certain essential properties. Can I coherently suppose that Shakespeare—the man himself—had different parents from the ones he in fact had, or that he was born in a different century? Kripke would argue that if we think clearly about what we can suppose these things, we are confused: if we think clearly about what we are trying to suppose we will see that these are not coherent counterfactual possibilities. Not everyone will share these intuitions, even after setting aside the bad reasons for resisting them that Kripke points out. The possible worlds framework does not settle such metaphysical questions, or even tell us how they should be settled; its job is to raise them, and to make clearer what the alternative answers say.

I have argued that Kripke's metaphysical theses do not presuppose his theses on reference and intentionality. What about the other direction? Does the Millian semantics for proper names or the causal account of reference and intentionality presuppose the metaphysical picture that Kripke defends? The theses about names can easily be separated from the more specific essentialist theses, but the general metaphysical issues about the identification of individuals across possible worlds are more difficult to disentangle from the thesis that names are rigid designators whose reference is established by causal interaction between the speaker and the referent.

[23] Searle (1969: 93).

[24] Wittgenstein (1953: §126–8).

A rigid designator is a designator that denotes the same individual in all possible worlds; doesn't this presuppose that the same individuals can be found in different possible worlds?

Consider the following anti-essentialist metaphysical picture, a version of the 'bundle of qualities' conception of an individual rejected by both Kripke and Searle. According to this conception, a particular individual is just the coinstantiation of a certain set of qualities. If individuals are identified across possible worlds at all, it is only in virtue of some counterpart relation which is definable in terms of the relations between the bundles of qualities found in the different worlds. Consider first the pure Leibnizian version of this metaphysical picture according to which particular individuals have all of their properties, including their relational properties, essentially. On this conception, not only is it a mistake to think that we can coherently suppose that Shakespeare—the person himself—had different parents, we cannot even make sense of a counterfactual possibility in which he ate a slightly different breakfast than he in fact ate on a certain morning, or even lived in a world in which slightly different events took place years after his death. The thesis that names are rigid designators is perfectly compatible with this uncompromising metaphysics, but the combination of metaphysical and semantic theses has no plausibility. It gives us the conclusion, for example, that the proposition expressed by the statement that Shakespeare wrote plays is a proposition that is true only in the actual world, and so is a proposition that entails every true proposition. Only God could know that Shakespeare wrote plays. We more limited creatures can know that 'Shakespeare wrote plays' expresses a true proposition, and we can know that whatever proposition it expresses, it is necessarily equivalent to the proposition expressed by 'Elvis Presley played the guitar', but so long as we are ignorant of any fact, we cannot know which proposition it is that these sentences express.

Giving up the Millian theory of names would not resolve the problem, since there is no plausibility in the assumption that however we refer to him, Shakespeare—the man himself—could not possibly have failed to write plays, or to eat what he in fact ate for breakfast. A less uncompromising version of this metaphysical picture gives a different account of what it is for an individual to have a property essentially. According to the liberal Leibnizian,[25] to say that Shakespeare—the person himself—might not have written plays is to say that a counterpart of that person in some possible world did not write plays, where the counterpart relation is reducible to some kind of qualitative similarity. The counterpart variation of the Leibnizian metaphysics of individuals is not a thesis about names—it is about the modal properties of individuals, however they are referred to. The difference between the unreconstructed Leibnizian theory and the

[25] Lewis (1986a: ch. 4).

counterpart version might still be construed as a semantic rather than a metaphysical difference, but it is a difference in the way complex predicates involving modality are to be interpreted, and by itself says nothing about how names are to be understood. But the absurd consequences drawn from the combination of the Leibnizian metaphysics and the Millian account of names were about the propositions expressed by simple sentences involving nonmodal predicates. If 'Shakespeare' is a Millian name, then 'Shakespeare wrote plays' is true in a possible world only if Shakespeare himself wrote plays in that world, which means only in the actual world. The fact that we construe 'could have written no plays' in such a way that 'Shakespeare could have written no plays' is also true in (and only in) the actual world is beside the point.

One might try to reconcile the Leibnizian metaphysics in its counterpart variation with a version of the Millian semantics for names by reinterpreting the concept of a rigid designator in a way that parallels the reinterpretation of modal predication. Suppose we say that a designator is *quasi-rigid* (relative to a possible world w) if its referent in any possible world w' is a counterpart of some particular individual is w. Then if 'Shakespeare' is a quasi-rigid designator, relative to the actual world, 'Shakespeare wrote plays' might be true in some other possible worlds—worlds in which a counterpart of Shakespeare wrote plays. The thesis that names are quasi-rigid designators is a kind of compromise between the Millian theory and the sense theory that accounts for some of the phenomena Kripke brought to our attention, but there are problems with it. First, if (as David Lewis has argued) an individual may have more than one counterpart in the same world, then the semantic value of a name will not be determined by the individual. Suppose Shakespeare has two counterparts in some possible world, only one of whom wrote plays. Is 'Shakespeare wrote plays' true in that possible world or not? To answer this we need to know which of two (or more) quasi-rigid concepts is expressed by 'Shakespeare'. Second, even if an individual has at most one counterpart in any possible world, a designator might be quasi-rigid relative to one possible world, and not relative to another, since two counterparts of the same thing (in different possible worlds) need not be counterparts of each other.

Despite these problems, the basic Kripkean picture of the way the reference of names is determined might still be reconciled with this metaphysical theory. What is essential to Kripke's picture, I think, is the idea that the content of speech acts and mental attitudes may be determined as a function of particular things (and kinds) with which the speakers and thinkers interact. Whatever one's metaphysical presuppositions, it will be agreed that the way content and reference are determined by the facts will be context-dependent, and influenced by general beliefs, purposes, and assumptions. The counterpart theorist's story about the background against which reference to a particular (world-bound) individual can determine a proposition

will be different from the story Kripke might tell, but there is nothing in the counterpart theorist's metaphysics that prevents him telling such a story.

The dialectic of this last discussion shows that metaphysical and semantic issues cannot be kept completely apart for at least two reasons. First, if semantic and metaphysical theses together yield implausible consequences, it may be a matter of dispute where the source of the problem lies. Second, some metaphysical theories may force a reformulation of claims about semantics and intentionality. But the possible worlds framework helps to clarify the metaphysical alternatives, and to separate metaphysical from semantic issues so that each can be evaluated on its own terms.

Ways a World Might Be

METAPHYSICAL AND ANTI-METAPHYSICAL ESSAYS

ROBERT C. STALNAKER

Clarendon Press · Oxford   2003