15-Jackson-Chap-15.qxd 17/5/05 5:59 PM Page 395

PARTIV

PHILOSOPHY OF LANGUAGE

15-Jackson-Chap-15.qxd 17/5/05 5:59 PM Page 396

Œ

-(

CHAPTER 15

REFERENCE AND DESCRIPTION

SCOTT SOAMES

1. THE REVOLT AGAINST DESCRIPTIVISM

1.1 Names

The modern discussion of reference begins with a revolt, led by Saul Kripke, against theories of the meaning and reference of proper names inspired by Gottlob Frege (1970) and Bertrand Russell (1905). In 'Naming and Necessity' (1980) Kripke attacked both the view that the meanings (semantic contents) of names are given by descriptions associated with them by speakers, and the view that their referents are determined (as a matter of linguistic rule) to be the objects that satisfy such descriptions. The view about reference is taken to follow from the view about meaning, but not vice versa. Thus, all of Kripke's arguments against descriptive theories of reference are also arguments against descriptive theories of meaning, but some arguments against the latter do not apply to the former.

We begin with these arguments. Let n be a proper name, D be a description or family of descriptions associated with n by speakers, and $\dots D^* \dots$ be a sentence that arises from $\dots n \dots$ by replacing one or more occurrences of n with D^{*}. When D is a description we let $D^* = D$, when D is a family of descriptions $D_1 \dots D_k$, we let D^* be the complex description *the thing of which most, or a sufficient number, of the claims: it is D*₁...*it is D*_k *are true*. Kripke attacks the following corollaries of descriptivity about the meanings of names.

- C. Since the semantic content of (proposition expressed by)...n... (as used by s in context C) is the semantic content of (proposition expressed by)...D*... (as used by s in C),
 - M. ... n... is true w.r.t. s, C, and a possible world-state w iff... D^* ... is true w.r.t. C, and w. Since
 - (i) If D^* exists, then D^* is D^*

is a necessary truth,

(ii) If *n* exists, then *n* is D^*

is also necessary;

E. anyone who knows/believes the proposition expressed by $\dots n \dots$ (w.r.t. C) knows/believes the proposition expressed by $\dots D^* \dots$ (w.r.t. C), and the ascriptions **Ralph knows/believes that n is F** and **Ralph knows/believes that D* is F** (as used by s in C) agree in truth-value. Since the proposition expressed by (ii) is the same as the proposition expressed by (i), it is knowable a priori, and the claim, **It is knowable a priori that if n exists, then n is D*** is true.

Kripke's argument against M is known as *the modal argument*. Here is a particular version of it. Consider the name *Aristotle*, and the descriptions *the greatest student* of *Plato*, *the founder of formal logic*, and *the teacher of Alexander the Great* that many of us associate with it. Although Aristotle satisfies these descriptions,

(1) If Aristotle existed, then Aristotle was D^*

is not a necessary truth, where D^* is either any description in this family, or the complicated description *the individual of whom most, or a sufficient number, of the claims:...are true* constructed from the family. On the contrary, Aristotle could have existed without doing any of the things for which he is known; he could have moved to another city as a child, failed to go into philosophy, and never been heard from again. When evaluated at such a possible scenario the antecedent of (1) is true, while the consequent is false. Since (1) is false in this scenario, it is not necessary truth, which means that this family of descriptions doesn't give the meaning of *Aristotle*. According to Kripke, this is no accident; there is, he suggests, no family D_A of descriptions such that (i) the referent of *Aristotle* is the unique individual who satisfies most, or a sufficient number, of the descriptions in D_A , (ii) ordinary speakers associate D_A with *Aristotle*, believing its referent to be the unique individual who satisfies most, or a sufficient number, of D_A 's members, and (iii) (1) expresses a necessary truth when D^* is the complicated description constructed from D_A . If this is right, then M is false, and names are not synonymous with descriptions associated with them by speakers.

The key point in the argument is Kripke's conviction that there was an individual x (Aristotle) such that a sentence *Aristotle was F* is true at an arbitrary world-state

w iff at w, x had the property expressed by F. Otherwise put, there was a unique individual x such that for any predicate F and world-state w, the proposition we actually use *Aristotle was F* to express would have been true if the world were in w iff were the world in w, x would have had the property (actually) expressed by F. This is the basis of his doctrine that *Aristotle* is a **rigid designator**.

Definition of rigidity. A singular term t is a rigid designator of an object o iff t designates o in all worlds in which o exists, and t never designates anything else.

Intuitive test for rigidity. A singular term t is a rigid designator iff *the individual who is t could not have existed without being t, and no one who is not the individual who is t could have been t* is true; otherwise t is non-rigid.

Using the notion of rigidity, we can state the general form of the modal argument.

The general version of the modal argument

- (i) Proper names are rigid designators.
- (ii) If a description D gives the meaning (content) of a term t, then D is rigid iff t is.
- (iii) So, the meanings or contents of proper names are not given by non-rigid descriptions.

Since the descriptions we have been considering are non-rigid, the meaning of *Aristotle* is not given by them. In *Naming and Necessity*, Kripke leaves the modal argument there, concluding that there are no meaning-giving descriptions associated with names by speakers. In so doing, he appears, tacitly, to assume that the only candidates for being meaning-giving descriptions are non-rigid. Though understandable, this assumption is not beyond question, and will be revisited later. Kripke's epistemological argument against E may be reconstructed as follows:

The epistemological argument

- (i) When D* is a description concerning well-known characteristics of the referent of an ordinary name n, it is not the case (i) that one knows or believes the proposition expressed by *n* is *F* iff one knows or believes the proposition expressed by *D** is *F*, (ii) that *Ralph knows/believes that n is F* and *Ralph knows/believes that D** is *F* invariably agree in truth-value, (iii) that the proposition expressed by *If n exists, then n is D** is knowable a priori, or (iv) that *It is knowable a priori that if n exists, then n is D** is true.
- (ii) So, descriptions concerning well-known characteristics of the referents of ordinary names do not give their meanings (semantic contents).
- (iii) Since these are the descriptions standardly associated with names by speakers, E is false, and the meanings of names are not given by descriptions speakers associate with them.

Two features of the argument stand out. First, it tacitly assumes that if n meant the same as D^* , then **Ralph believes, knows, or knows a priori that n is F** would have the same truth-value as **Ralph believes, knows, or knows a priori that** D^* **is F**. Although a case can be made that Kripke did tacitly assume this, the theoretical basis for this assumption goes beyond what he explicitly stated. Secondly, the argument form is general, and need not be limited to claims about a priori knowledge, or even to claims about knowledge and belief, as opposed to other attitudes. The point is to show that the proposition expressed by *n* is *F* is different from the proposition expressed by *D* is *F*. This is done by showing that a person can bear a certain attitude to one of these propositions without bearing it to the other. Although the relation one bears to a proposition to which one might appeal.

We now turn to Kripke's reason for accepting premiss (i). His text is replete with thought experiments supporting it, one of them being the famous Gödel-Schmidt example, concerning the origins of Gödel's famous incompleteness theorem (Kripke 1980: 83-4). Kripke imagines our belief that Gödel discovered the incompleteness theorem being proven false by historical scholarship that reveals that he stole it from Schmidt. Of course, Kripke is not saying that any such thing really happened, or even that we don't know that it didn't. The point is that we don't know this a priori. By contrast, we do know a priori that the discoverer of the incompleteness of arithmetic discovered the incompleteness of arithmetic (if anyone did). So, the proposition expressed by Gödel discovered the incompleteness of arithmetic (if anyone did) is not the same as the proposition expressed by The discoverer of the incompleteness of arithmetic discovered the incompleteness of arithmetic (if anyone did), and the description does not give the meaning of the name Gödel for us-no matter how central attribution of the theorem to Gödel is to our beliefs about him. Kripke takes this point to extend to other descriptions speakers may associate with Gödel, and to proper names generally. He concludes that the meanings of proper names are not given by any descriptions that pick out an individual in terms of famous achievements, or important characteristics. Given this, one is hard-pressed to see how they could be given by any descriptions at all. Kripke therefore concludes that E is false, and that names are not synonymous with descriptions.

He is, however, careful to distinguish this conclusion from one that holds that the referents of proper names are not determined, as a matter of linguistic rule, to be whatever objects satisfy the descriptions associated with them by speakers. According to this weakened version of descriptivism, descriptions associated with a name semantically fix its referent at the actual world-state, without giving its meaning. Once its referent is determined there, it is stipulated to retain that referent with respect to all other world-states; thus it is a rigid designator. Several corollaries are taken to follow. (i) The speaker has a description, or family of descriptions, D associated with n that he takes to be uniquely satisfied by some object or other. (ii) An object o is the referent of n iff o uniquely satisfies D (or a sufficient number of the descriptions in D, if D is a family). (iii) Since the speaker knows this on the basis of semantic knowledge alone, the speaker knows on that basis that the sentence If n exists, then n is D^* expresses a truth. In sum, when D semantically fixes the reference of n, understanding n requires knowing that its reference is fixed by D. This holds even though D does not give the meaning of n.

Kripke's arguments against this version of descriptivism are known as *the semantic arguments*, which are designed to constitute counter-examples to each of its corollaries. The Gödel–Schmidt scenario is taken to provide a counter-example to both (ii) and (iii). It is a counter-example to (iii) because we don't know, simply on the basis of our linguistic competence, that the sentence *If Gödel existed, then Gödel was D* is true, when D is a description, or family of descriptions encompassing our most important knowledge of Gödel. It is a counter-example to (ii) because when one imagines a state of the world just like ours except that, unknown to speakers, Kripke's fantasy about Gödel's plagiarism is true, we take those speakers to be referring to Gödel, not Schmidt, when they use the name *Gödel*. Thus, the referent of *Gödel*, as used by those speakers, is not the individual that satisfies the descriptions they associate with it. If these arguments are correct, then description theories of the referents of names are incorrect.

There is, however, a distinction to be made. Although Kripke suggests that the meanings of names are never given by descriptions that speakers associate with them, he does allow that in rare cases the referent of a name may be semantically fixed by a description. In most normal cases, however, proper names are seen as having their referents fixed by a historical chain of reference transmission. Typically, the chain begins with an ostensive baptism in which an individual is stipulated to the bearer of a name n. Later, when n is used in conversation, new speakers encounter it for the first time and form the intention to use it with the same reference as their sources. Different speakers may, of course, come to associate different descriptions with n, but in the usual case this doesn't affect reference transmission. As a result, speakers further down the historical chain may use n to refer to its original referent o, whether or not they associate descriptions with n that uniquely denote o.

So Kripke has an alternative to descriptivist theories of reference determination. What about meaning? On his account, it would seem that the only semantic function of a name is to refer, in which case one would expect ordinary proper names to be regarded as Russellian logically proper names (without Russell's epistemological restrictions on their bearers) (for discussion of Russell's doctrine of logically proper names, see Soames 2003*a*, ch. 5). However, Kripke does not draw this, or any other, positive conclusion about the meanings of names, or the propositions semantically expressed by sentences containing them. Along with nearly everyone else, he recognizes that one can understand different co-referential names without knowing that they are co-referential, and certainly without judging them to have

the same meaning. However, this doesn't show that they don't mean the same thing, unless one accepts the questionable principle that anyone who understands a pair of synonymous expressions must recognize them to be synonymous (or at least coextensive)—something Kripke never does. In Naming and Necessity, he does argue that Hesperus is Phosphorus is not knowable a priori, whereas the triviality that Hesperus is Hesperus is (for explication and criticism of Kripke's argument, see Soames 2003B, ch. 15). This view, together with natural assumptions about meaning, compositionality, propositions, and propositional attitude ascriptions, could be used to argue that Hesperus and Phosphorus differ in meaning, despite being co-referential. However, Kripke neither gives such an argument, nor draws such a conclusion. Moreover, he has no account of what, over and above their referents, the meanings of names might be. Finally, in 'A Puzzle About Belief' (1979) he maintains that no definite conclusions can be drawn about the meanings of names from apparent failures of substitution of co-referential names in belief ascriptions. On his view, these puzzles arise from principles of belief attribution that transcend any doctrine about the meaning of names. Hence, he resists drawing any positive conclusion about their meaning, or about the propositions semantically expressed by sentences containing them.

1.2 Natural Kind Terms

The challenge to descriptivism is not limited to proper names, having been pressed by Kripke (1980, lect. 3) and Hilary Putnam (1970, 1973, 1975) against descriptive analyses of natural kind terms like gold, tiger, water, heat, colour, and red. Kripke argues that, like proper names, these terms are not synonymous with descriptions associated with them by speakers; and, like names, they acquire reference in two ways. One way involves direct presentation of samples, together with the stipulation that the term is to apply to all and only instances of the unique natural kind (of a certain sort) of which nearly all members of the sample are instances; the other involves the use of a description to pick out a kind by some, usually contingent, properties. Later, when the term is passed from speaker to speaker, the way in which the reference was initially established normally doesn't matter-just as with names. As a result, speakers further down the linguistic chain may use the term to apply to instances of the kind, whether or not the descriptive properties they associate with the term really pick out its members. In addition, scientific investigation may lead to the discovery of properties that are necessary and/or sufficient for membership in the kind. According to Kripke, these discoveries are expressed in theoretical identification sentences like those in (2), which are necessary, but knowable only a posteriori.

(2*a*) Water is H_2O .

(2*b*) Lightning is electricity.

1.3 Indexicals, Quantification, and Direct Reference

15-Jackson-Chap-15.gxd

17/5/05

Starting with lectures given by David Kaplan in 1971, and continuing with published work of Kaplan (1979, 1989) and John Perry (1977, 1979), a further challenge to descriptivism was mounted, focusing on the role of context in understanding indexicals like *I*, *now*, *today*, *here*, *actually*, *you*, *she*, and *that*. Although the referents of these terms vary from one context of utterance to another, their meanings do not. For example, to know the meanings of *I*, *today*, and *she* is, roughly, to know the rules in (3).

- (3*a*) One who uses *I*—e.g. in a sentence *I* am *F*—refers to oneself, and says of oneself that one 'is F'.
- (3*b*) One who uses *today*—e.g. in a sentence *Today is F*—refers to the day the utterance takes place, and says of that day that it 'is F'.
- (3*c*) One who uses *she*—e.g. in a sentence *She is F*—refers to a contextually salient female, and says of her that she 'is F'.

These rules provide two kinds of information: they tell us how the referents of indexicals depend on aspects of contexts in which they are used; and they implicitly identify the semantic contents of these terms with their referents in contexts.

In order to understand this talk of content, one must grasp Kaplan's intuitive semantic framework. Sentences express propositions, which are their semantic contents; those containing indexicals express different propositions, and so have different contents, in different contexts. Nevertheless, the meaning of a sentence is constant; it is a function from contexts to contents. Kaplan's word for this is *character*. The picture is recapitulated for subsentential expressions. For example, the character of *I* is a function that maps an arbitrary context *C* onto the agent (speaker) of C, which is the semantic content of *I* relative to C.

There are two anti-descriptivist implications here. First, the referents of at least some indexicals are not determined by descriptions speakers associate with them. One example from Perry (1977: 487) is Rip Van Winkle, who awakes on 20 October 1823 after sleeping for twenty years, and says, not realizing what happened, *Today is 20 October 1803*. In so doing, he speaks falsely because his use of *today* refers to the day of the context, no matter what description he may have in mind. Another example involves Kaplan's identical twins Castor and Pollux, raised in qualitatively identical environments to be molecule for molecule identical and so, presumably, to associate the same purely qualitative descriptions with the same terms (Kaplan 1989: 531). Despite this, each refers to himself, and not the other, when he uses *I*. These examples show that the referents of indexicals are not always determined by purely qualitative descriptions that speakers associate with them. Although this leaves open the possibility that some indexicals may have their referents semantically fixed by descriptions containing other indexicals, it precludes the possibility that all indexical reference is determined in this way.

The second anti-descriptivist implication is that since the semantic content of an indexical in a context is its referent, its content is not that of any description. In order make this point, one must move beyond the formal system developed in 'On the Logic of Demonstratives' (Kaplan 1979), to the conception of structured contents that Kaplan characterizes in 'Demonstratives' (1989:496-7) as the intuitive philosophical picture underlying his approach. On this picture, the proposition expressed by S in C is a complex encoding the syntactic structure of S, the constituents of which are (or encode) the semantic contents in C of the words and phrases in S. For example, the proposition expressed in C by a sentence *i* is *F* is a complex in which the property expressed by F is predicated of the referent o of the indexical i; this is the same as the singular proposition expressed by x is F, relative to an assignment of o to the variable "x". By contrast, the proposition expressed by *The* **D** is F in C is a complex consisting of the property expressed by F plus a complex consisting of (or encoding) the content of *the* together with the structured complex which is the semantic content in C of the descriptive phrase D. On one natural analysis, this proposition ascribes the higher-order property of being instantiated by whatever uniquely instantiates the property expressed by D to the property expressed by F.

The claim that the semantic content of an indexical relative to a context is not the same as that of any description is supported by commonplace observations about propositional attitudes. Suppose, to adapt Russell's famous example, that on some occasion in which George IV spied Walter Scott, he gave voice to his newfound conviction, saying *He* [gesturing at Scott] isn't the author of Waverley. Had this occurred, each of the following ascriptions would have been true.

- (4*a*) The author of *Waverley*, namely Scott, is such that George IV said that he wasn't the author of *Waverley*.
- (4*b*) George IV said that you weren't the author of *Waverley*. (said addressing Scott)
- (4*c*) George IV said that I wasn't the author of *Waverley*. (said by Scott)
- (4*d*) George IV said that he [pointing at Scott] wasn't the author of *Waverley*. (said by a third party in another context)

On Kaplan's picture, these reports are true because the semantic content of the sentence George IV uttered (in his context), and so the proposition he asserted, is the same as the content of the complement clauses in the reports of what he said. Whatever descriptions speakers who utter (4b,c,d) may associate with the indexicals are irrelevant to the semantic contents of the reports.

We are now ready to define the notion of a directly referential term, and contrast it with a relativized notion of rigid designation.

Direct reference. A term t is directly referential iff for all contexts C and assignments A, the referent of t with respect to C and A = the content of t with respect to C and A.

Relativized rigid designation. A singular term t is a rigid designator iff for all contexts C, assignments A, world-states w, and objects o, if t refers to o with respect to C, A, and w, then t refers to o with respect to C, A, and w in all world-states w in which o exists, and t never refers to anything else with respect to C, A, and any world-state w^{*}.

With this understanding, all directly referential singular terms are rigid designators, but not vice versa (e.g. *the square root of 25* is rigid but not directly referential). According to Kaplan, indexicals and variables are directly referential. Salmon (1986) and Soames (2002) extend this view to proper names.

Before leaving Kaplan's framework it is important to consider two indexical operators used to construct rigidified descriptions out of non-rigid descriptions. One, *dthat*, combines with a description D to form a directly referential singular term *dthat* [D] the content of which, relative to C and A, is the unique object denoted by D relative to C and A. The other, *actually*, stands for the world-state C_w of the context in a manner analogous to the way in which now stands for the time C_v and I stands for the agent C_A , of the context. *Actually* combines with a sentence S to form a complex sentence *Actually S* the content of which in C is a proposition that predicates of C_w the property of being a world-state in which the proposition expressed by S in C is true; hence *Actually S* is true with respect to C and world-state w iff S is true with respect to C and C_w , and whenever S is true in C_w , *Actually S* is a necessary truth. The corresponding fact about descriptions is that whenever *the x: Fx* denotes a unique individual o in C_w , *the x: actually Fx* denotes o with respect to C and all possible world-states in which o exists, and never denotes anything else. Hence actually is a rigidifier; however, the resulting rigidified descriptions are not directly referential.

1.4 Philosophical Implications

15-Jackson-Chap-15.qxd

17/5/05

Although it seems evident that the propositions expressed by (5) are knowable only a posteriori, it appears to be a consequence of the non-descriptive semantics of names, natural kind terms, and the actuality operator that each of these sentences expresses a necessary truth, if it is true at all.

- (5*a*) Saul Kripke \neq David Kaplan.
- (5*b*) Lightning is electricity.
- (5*c*) If Thomas Jefferson existed, then the person who actually wrote the Declaration of Independence was Thomas Jefferson.

Since we know that (5a) - (sc) are true, it follows that they are examples of the necessary a posteriori. The best examples of the contingent a priori are sentences like (6).

(6*a*) If someone wrote the Declaration of Independence, then the person who actually wrote the Declaration of Independence wrote something.

It follows from the semantics of *actually* plus the fact that Thomas Jefferson wrote the Declaration of Independence that (6) is false in a world-state in which Thomas Jefferson wrote nothing and someone else wrote the Declaration of Independence. Assuming that the world could have been in such a state, we conclude that (6) expresses a contingent truth. However, this truth can be known without doing empirical research. Since anyone in the actual world-state can know it to be true simply by understanding it, it would seem to be knowable a priori. On this view, then, not all necessary truths are a priori, and not all a priori truths are necessary.

2. An Attempt to Revive Descriptivism

2.1 Motivations

Despite the attack on descriptivism, some believe that the anti-descriptivists' conclusions are too extreme, and that properly modified descriptive analyses are capable of withstanding their arguments. This view is fuelled by three main factors. First is the conviction that anti-descriptivists have not adequately addressed Frege's puzzle about substitution of co-referential terms in attitude ascriptions and Russell's problem of negative existentials. There is still a widespread belief that these problems show that names cannot be directly referential. Although Kripke never asserted that they were, it is hard to see how, if his doctrines are correct, they could be anything else. According to him, the meaning of a name is never the same as that of any description, and the vast majority of names do not even have their referents semantically fixed by descriptions. If these names are so thoroughly non-descriptional, it is not clear how their meanings could be other than their referents. Consequently, one who takes that view to have been refuted by Frege and Russell will suspect that the power of Kripke's arguments must have been exaggerated, and will be motivated to find a way of modifying descriptivism that can withstand them.

The second factor motivating descriptivists is their conviction that critics like Kripke have focused on the wrong descriptions. To be sure, it will be admitted, for many speakers s and proper names or natural kind terms n, the descriptions most likely to be volunteered by s in answer to the question *To what, or to whom, do you refer when you use n*? neither give the meaning of n, nor semantically fix its reference. Often s will respond by citing what s takes to be the most well-known and important characteristics of the putative referent(s), about which s may be mistaken. However, the referents of these terms must be determined in some way, and surely, whatever way it is can be described. Thus, for each name or natural kind term n, there must be some description D that correctly picks out its referent(s)—perhaps one encapsulating Kripke's own causal-historical picture of reference transmission.

Is there any reason to believe that speakers associate D with n? Some descriptivists think so. In fact, the very success of Kripke and others in eliciting uncontroversial judgements about what names would refer to if used in various counterfactual situations has been taken to show that speakers must be implicitly guided by a descriptive theory that determines reference. For example, Frank Jackson (1998*b*: 212) argues that

if speakers can say what refers to what when various possible worlds are described to them, description theorists can identify the property associated in their minds with, for example, the word 'water': it is the disjunction of the properties that guide the speakers in each particular possible world when they say which stuff, if any, in each world counts as water. This disjunction is in their minds in the sense that they can deliver the answer for each possible world when it is described in sufficient detail, but it is implicit in the sense that the pattern that brings the various disjuncts together as part of the, possibly highly complex, disjunction may be one they cannot state.

This is a remarkable defence. If correct, it would mean that descriptive theories of reference are virtually guaranteed, a priori, to be irrefutable, since any refutation would require a clear, uncontroversial description of a possible scenario in which n referred to something o not satisfying a description putatively associated with n by speakers—whereas the very judgement that n does refer to o in this scenario would be taken by Jackson to demonstrate the existence of an implicit description in our minds that successfully determines reference, whether or not we can articulate it.

The third factor motivating a descriptivist revival involves the inability of some to see how any single proposition could be either both necessary and a posteriori, or both contingent and a priori, as anti-descriptivists maintain. How, these philosophers ask, can evidence about the actual world-state be required to establish p, if p is true in every possible state in which the world could be (including states in which no such evidence exists)? Or again, if q is contingent, then there are states that the world could be in, such that were the world in them, q would be false. But how could one possibly know without appeal to evidence that the world is not in such a state? The former worry casts doubt on the necessary a posteriori, the latter on the contingent a priori.

In addition to endorsing these sceptical worries, some descriptivists—e.g. Frank Jackson, David Chalmers, and David Lewis—adhere to antecedent philosophical commitments that make the existence of propositions that are both necessary and knowable only a posteriori impossible. One of these commitments is to metaphysical possibility as the only kind of possibility. Although these theorists recognize

different metaphysically possible ways that the world could be, they reject the idea that, in addition to these, there are epistemologically possible ways that the world might be-states which, though metaphysically impossible, cannot be known by us a priori not to obtain (Chalmers 1996:136–8; Jackson 1998a:67–74). This restriction of epistemic possibility to metaphysical possibility renders the necessary a posteriori problematic from the start—since it precludes explaining it by citing metaphysically necessary propositions for which empirical evidence is needed to rule out metaphysically impossible, but epistemologically possible, world-states in which they are false (for an explanation along these lines, see Soames 2005), forth coming When one adds to this Lewis's (1996: 422) analysis of knowing p as having evidence that rules out all possible ways of p's being untrue, one has, in effect, defined propositions that are both necessary and knowable only a posteriori out of existence. Since, when p is a necessary truth, there are no metaphysically possible ways of its being untrue, it follows that there are no possible ways of p's being untrue at all, and hence that knowledge of p never requires empirical evidence. So, the necessary a posteriori is impossible. A different philosophical commitment that leads to the same conclusion identifies propositions with sets of metaphysically possible world-states (or functions from world-states to truth-values). On this view, there is only one necessary proposition-which is known a priori (see Stalnaker 1978, 1984; Lewis 1996: 422–3; Jackson 1998a: 71–4, 75–8). But if that is right, then the anti-descriptivist semantics that leads to the conclusion that there are necessary a posteriori propositions must be mistaken.

2.2 Strategy

The main strategy for constructing descriptive analyses of names and natural kind terms combines attempts (i) to find reference-fixing descriptions capable of withstanding Kripke's semantic arguments, (ii) to avoid the modal argument, either by rigidifying these descriptions, or by insisting that they take widescope over modal operators in the same sentence, and (iii) to use two-dimensional semantics to avoid the epistemological argument, and explain away putative examples of the necessary a posteriori and the contingent a priori. The most popular strategy for finding reference-fixing descriptions is *causal descriptivism*, which involves extracting a description from Kripke's causal–historical account of reference transmission. Details aside, the general idea is clear enough, as is illustrated by the following passage from David Lewis:

Did not Kripke and his allies refute the description theory of reference, at least for names of people and places? Then why should we expect descriptivism to work any better for names of colors and color experiences? . . . I disagree. What was well and truly refuted was a version of descriptivism in which the descriptive senses were supposed to be a matter of famous deeds and other distinctive peculiarities. A better version survives the attack: *causal*

REFERENCE AND DESCRIPTION

409

descriptivism. The descriptive sense associated with a name might for instance be *the place I have heard of under the name 'Taromeo'*, or may be *the causal source of this token: 'Taromeo'*, and for an account of the relation being invoked here, just consult the writings of causal theorists of reference. (Lewis 1997. n. 22; see also 1984; Kroon 1987; Searle 1983, ch. 9; Chalmers 2002)

The second part of the descriptivists' strategy is the attempt to avoid Kripke's modal argument. The simplest way of doing this, once one has what one takes to be a reference-fixing description, is to rigidify it using *actually* or *dthat* (this strategy is advocated by Frank Jackson (e.g. 1998*b*: 213–14). An alternative method is to analyse names as meaning the same as non-rigid descriptions that are required to take large scope over modal operators (and modal predicates) in the same sentence, while taking small scope when embedded under verbs of propositional attitude (this strategy originated with Michael Dummett 1973). Although the details can be complicated, the guiding idea is simple. Descriptivists want to explain apparent instances of substitution failure involving co-referential proper names in attitude ascriptions by appealing to descriptive semantic contents of names occurring in the complement clauses; however, they also want to guarantee substitution success when one co-referential name is substituted for another in modal construction s. The different strategies of dealing with Kripke's modal argument are designed to do that.

The final weapon in the descriptivists' strategic arsenal is (ambitious) twodimensionalism. To illustrate, we begin with a pair of contingently co-designative descriptions that have been rigidified.

(7) If there is a unique object that is actually F and a unique object that is actually G, then the x: actually Fx = the y: actually Gy.

Since the descriptions are co-designative, (7) is a necessary truth. But if asked, *What does one need to know in order for (9) to be true of one?*, it is natural to answer that it is necessary and sufficient that one know that which is expressed by (8).

- (8) If there is a unique object that is F and a unique object that is G, then the x: Fx = the x: Gx.
- (9) x knows that if there is a unique object that is actually F and a unique object that is actually G, then the x: actually Fx = the x: actually Gx.

Since (8) is knowable only a posteriori, the two-dimensionalist concludes that (7) is an example of the necessary a posteriori.

A similar story can be told about the contingent a priori. Let *the x: Fx* be a nonrigid description which designates some individual o in the actual world-state, but which designates other individuals with respect to other possible world-states. (10) is then a contingent truth that is false with respect to world-states in which *the x: Fx* designates some individual other than o.

(10) If there is a unique F, then the x: actually Fx = the x: Fx.

But if one asks *What does one need to know in order for (12) to be true of one?*, the natural answer is that for this to be true, it is sufficient that one know that which is expressed by (11).

- (11) If there is a unique F, then the x:Fx = the x:Fx.
- (12) x knows that if there is a unique F, then the x:actually Fx = the x:Fx.

Since (11) is knowable a priori, the two-dimensionalist concludes that (10) is an example of the contingent a priori.

There are two different, and competing, lessons a two-dimensionalist might draw from these examples. One is that although the proposition expressed by (7) is genuinely both necessary and knowable only a posteriori, the reason it is a posteriori is that it has the peculiarity that it can be known only by virtue of knowing a different, related proposition (expressed by (8)); similarly, although the proposition expressed by (10) is genuinely both contingent and knowable a priori, the reason it is a priori is that it can be known by virtue of knowing the distinct proposition expressed by (11). This, as we shall see, leads to a view I call weak two-dimensionalism. A different possible two-dimensionalist lesson is that what is required by the truth of (9) and (12) is not that that the agent know the propositions expressed by (7) and (10) at all, but simply that the agent know the propositions expressed by (8) and (11). Although the necessary proposition expressed by (7) is, on this view, knowable a priori, this is compatible with the fact that the knowledge reported by (9) is (and can only be) a posteriori. Similarly, although the contingent proposition expressed by (10) is knowable only a posteriori, this is compatible with the fact that the knowledge reported by (12) is a priori. This way of looking at things is characteristic of *classic*, or *strong*, *two-dimensionalism*.

Although both versions of two-dimensionalism allow the descriptivist to maintain that he has accounted for examples of the necessary a posteriori and the contingent a priori, strong two-dimensionalism has greatest appeal for theorists like Lewis, Jackson, and Chalmers. In addition to smoothly accommodating their antecedent philosophical commitments, strong two-dimensionalism fits better than weak two-dimensionalism does with the causal-descriptive analyses of names and natural kind terms. If, in order to avoid Kripke's semantic arguments, one is going to analyse n along the lines of Lewis's the x: actually x is the thing *I have heard of under the name n*, then one will not want the result that whenever r reports the knowledge of an agent using an ascription, *a knows that n is so and* so, r's report can be true only if the agent knows that whatever individual the reporter r has actually heard of under the name n 'is so and so'. Strong twodimensionalism avoids this absurdity by analysing attitude ascriptions as reporting relations between the agent and certain propositions distinct from the propositions semantically expressed by their complement clauses. Since this is not true of weak two-dimensionalism, strong two-dimensionalism is, initially, the more attractive choice.

15-Jackson-Chap-15.gxd 17/5/05 5:59 PM Page 41

2.3 Strong Two-Dimensionalism

Although contemporary two-dimensionalism is often cast as a proposal for a systematic semantic theory of natural language, often what is provided is more of a theory sketch than a theory. The best-known formal two-dimensionalist model, given by Martin Davies and Lloyd Humberstone in 'Two Notions of Necessity' (1980), provides explicit semantics for a very simple language—a modal version of the propositional calculus with a normal necessity operator, an actuality operator, and a new operator that takes one outside the usual modal model in a certain way. Such models are triples—consisting of a set W of world-states (called 'worlds'), a designated actual world-state @, and a valuation function that assigns intensions (functions from world-states to extensions) to the non-logical vocabulary. The basic operators are:

Necessarily S is true in a model M w.r.t. w iff S is true in M w.r.t. all worlds in W. *AS* is true in M w.r.t. w iff S is true in M w.r.t. @.

FS is true in M w.r.t. w iff for all models M' that differ from M at most regarding which world @ ' is designated as actual, S is true in M' w.r.t. w.

A further operator, FA, is definable in terms of the last two.

FAS is true in M w.r.t. w iff for all models M' that differ from M at most regarding which world @' is designated as *actual*, S is true in M' w.r.t. @'.

According to Davies and Humberstone (1980: 3), '*FA* α says: whichever world had been actual, α would have been true at that world considered as actual'—not the most edifying formulation. A clearer explanation is that *FA* α says that α (understood as we actually understand it) expresses a truth in any context, no matter what the designated world-state of the context. They use this operator to capture Gareth Evans's (1979) distinction between *deep* and *superficial necessity*, which are translated into their system as follows:

A true sentance S is superficially contingent iff *Necessarily S* is false.

A true sentence S is deeply contingent iff FAS is false.

S is superficially necessary iff *Necessarily S* is true.

S is deeply necessary iff *FAS* is true.

Davies and Humberstone support Evans's idea that all examples of the contingent a priori are sentences that are superficially contingent, but deeply necessary, and that this provides an explanation of their status. But what, precisely, is the explanation? How is it that *It is a necessary truth that S* may be false when *It is knowable a priori that S* is true? Since Davies and Humberstone do not offer any fully precise theory about this, we need to supplement their account with explicit semantics for *it is knowable a priori*. Think of an *indexical model* as a family of Davies–Humberstone models, each of which differ from the others only in the

world-state designated as *actual*. Call these designated world-states *contexts*. Then *FA* is an operator on Kaplan-style characters, and the proposition expressed by *FAS* is the set of all contexts/world-states iff for every context C, the character of S assigns to C a proposition true at C; otherwise *FAS* expresses the empty set of contexts/world-states. The characteristic two-dimensionalist claim about the a priori is that S is a priori iff *FAS* is true.

One way to think about this would be to characterize *it is knowable a priori* — as operating on the character of S—even as being synonymous with *FA*.¹ Though this is a possible alternative, it is not, by itself, very revealing, since surely we understand the operator on the basis of understanding the two-place predicate — *knows apriori* — the second argument of which is the same as the second argument of — *knows* —. Regarding these predicates, the Davies–Humberstone picture allows two natural variations—to treat *knows* as a two-place predicate of an agent and the character of its complement sentence, or to treat it as a two-place predicate to which the character of its complement sentence assigns the value truth. The latter, having the advantage of making the object of knowledge a proposition, has proven to be the more popular.

We can now give the desired strong two-dimensionalist explanation of the contingent a priori. For any sentence S, the *primary intension* of S is the proposition that its character expresses a truth—i.e. the set of contexts/world-states C to which the character of S assigns a proposition true at C. The proposition expressed by S at C is the *secondary intension* of S at C. Although the primary and secondary intensions of some sentences will be the same, in many cases in which S contains the actuality operator—e.g. (7) and (10)—they will be different. *It is a necessary truth that S* is true in C iff the secondary intension of S in C is true in all possible world-states; *It is knowable a priori that S* is true in C iff the primary intension of S is knowable a priori. Davies and Humberstone suggest that in every genuine case of the contingent a priori, since its secondary intension is contingent while its primary intension—which is the secondary intension of (11)—is both necessary and a priori. There is no puzzle about how any one proposition can be both contingent and a priori, since none is.

The account easily extends to examples like (7). Although its secondary intension is necessary, its primary intension is identical with the contingent, a posteriori truth that is the secondary intension of (8). As before, there is no puzzle about the necessary a posteriori status of (7), since the proposition that is necessary is not the one that is knowable only a posteriori. Of course, we have so far considered only examples of the contingent a priori and the necessary a posteriori containing the actuality operator. To extend the view to all Kripkean instances of these two-categories, the

¹ Stalnaker (1978: 83) calls what is essentially this operator 'the apriori truth operator.'

two-dimensionalist must analyse all names and natural kind terms as rigidified descriptions (not containing any unanalysed names or natural kind terms).² This, in a nutshell, is descriptive strong two-dimensionalism. According to it, the necessary a posteriori and the contingent a priori are, in effect, linguistic illusions, born of equivocation between primary and secondary intension.

3. Critique of the Descriptivist Revival

3.1 Problems with Strong Two-Dimensionalism

The first difficulty with this attempt to revive descriptivism is that strong twodimensionalism faces powerful counter-arguments (additional arguments of the same sort can be found in Soames 2005*b*, forth coming).

Argument 1

15-Jackson-Chap-15.qxd

17/5/05

- According to strong two-dimensionalism, epistemic attitude ascriptions *a knows/believes that S* report that the agent bears the knowledge/belief relation to the primary intension of S.
- (2) Let the F and the G be two contingently co-designative, non-rigid descriptions. Since for every context C, the character of the F = the G expresses a truth with respect to C iff the character of the actual F = the actual G does too, the two primary intensions are identical. Thus, the ascriptions a knows/believes that the actual F = the actual G and a knows/believes that the F = the G are necessarily equivalent.
- (3) Hence, the truth-value of

Necessarily [if the actual F = the actual G and Mary believes that the actual F = the actual G, then Mary believes something true]

is the same as the truth-value of

Necessarily [if the actual F = the actual G and Mary believes that the F = the G, then Mary believes something true].

Since the latter modal sentence is false, so is the former.

² Davies and Humberstone express a healthy degree of scepticism about this part of the project. As a result, they cannot be seen as definitively embracing full-fledged strong, descriptive two-dimensionalism. Nevertheless, they provided the semantic basis for the later development of this view. For a full account of the details and nuances of their paper see Soames 2005*b*, ch. 5.

(4) Similarly, the truth-value of

Necessarily [if Mary believes that the actual F = the actual G, and if that belief is true, then the actual F = the actual G].

is the same as the truth-value of

Necessarily [if Mary believes that the F = the G, and if that belief is true, then the actual F = the actual G].

Since the latter modal sentence is false, so is the former.

(5) Since, in fact, the initial modal sentences in (3) and (4) are true, strong twodimensionalism is false; modal and epistemic operators in English do not take systematically different objects.

Argument 2

- (1) According to strong two-dimensionalism, names are synonymous with rigidified descriptions. Let o be an object uniquely denoted, in our present context, by the non-rigid description *the D*, let n be a name of o, and let the strong two-dimensionalist analysis of n be *the actual D*.
- (2) Then, according to strong two-dimensionalism, *John believes that n is the D* will be true (as used in our present context) relative to an arbitrary counter-factual world-state w iff *John believes that the D is the D* is true with respect to w. (I here assume (1) and (2) of argument 1.)
- (3) This means that in counterfactual world-states with respect to which *the D* doesn't denote o, and hence in which *the actual D = the D*, and *n = the D*, are false, *John believes that n is the D* will count as reporting the same belief, with the same truth-value, as *John believes that the D is D*. But surely, beliefs which are not about the referent of n can't correctly be described by sentences of the form *John believes that n*...
- (4) In general, for any agent a and any counterfactual world-state w, if a name n denotes an object o in our present context C, then (i) *John believes that n is F* (as used in C) is true when evaluated at w only if (ii) *John believes that y is F* is true at w with respect to an assignment of o to the variable "y". Strong two-dimensionalism is false, since it wrongly allows (i) to be true at w when (ii) is false at w.

Further arguments against strong two-dimensionalism make crucial use of indexicals like *I* and *now* the referents of which are not fixed by descriptions. In order to integrate these expressions into the two-dimensionalist framework, one must expand contexts to include designated agents and times, in addition to designated world-states. This has an immediate formal consequence. In the original Davies–Humberstone system, the meaning of S was a function from world-states, to propositions expressed by S relative to those world-states, and the primary intension of S was a proposition true when evaluated at w iff the meaning of S assigns w

a proposition that is true at w. Cleaving to the standard, strong two-dimensionalist identification of propositions with sets of world-states we may identify the primary intension of S with the set of world-states w, such that the meaning of S assigns to w a set of world-states containing w. Hence, the primary intension of S and the secondary intension of S (relative to a context) are both propositions (sets of world-states), and the roles of being a context in which sentences expresses propositions, and of being a counterfactual circumstance in which propositions are evaluated, are two different dimensions of the same thing—a world-state.

This changes when contexts (but not circumstances of evaluation) are expanded to include designated agents and times. Although the secondary intension of S relative to C remains a set of world-states, which we may call an *ordinary proposition*, the primary intension of S is now identified with the set of contexts to which the meaning of S assigns ordinary propositions that are true in (contain) the world-states of the contexts. Call these *pseudo-propositions*. We no longer have one thing which plays the roles of both context and circumstance of evaluation, and we no longer have one kind of thing that occurs as arguments to both modal operators and *it is knowable a priori* —. Instead, we have contexts and world-states in the first case, and ordinary propositions and pseudo-propositions in the second.

What is it to believe or know something which is not an ordinary proposition? Perhaps examples will help.

- (13*a*) I am here now.
- (13*b*) I am not Saul Kripke.

If I were to assertively utter (13a) now I would express the (ordinary) proposition that SS is in Princeton at 11.30 a.m. on 22 August 2002. Although this proposition is contingent, the meaning of (13a) generates a truth in every context; so strong twodimensionalism classifies it as a case of the contingent a priori. The situation is just the reverse with (13b), which is classified as necessary and a posteriori. Although these characterizations are defensible, they raise a question. What do I know in these cases, and what do I report myself as knowing, when, in the same context, I assertively utter (14a) and (14b)?

- (14*a*) I know that I am here now.
- (14*b*) I know that I am not Saul Kripke.

The natural answer is that what I know, and report myself as knowing, is the same as what I (truly) report Gideon as knowing when I assertively utter (15a)) and (15b)) and our new graduate student, Harold, as not knowing when I assertively utter (16a) and (16b).

- (15*a*) Gideon knows that I am here now.
- (15*b*) Gideon knows that I am not Saul Kripke.
- (16*a*) Harold doesn't know that I am here now.
- (16*b*) Harold doesn't know that I am not Saul Kripke.

What is that? Not the primary intensions of (13a) and (13a). Each of us accepts (in his own context) the meanings of (13a) and (13b) which in turn generate these primary intensions. (Each would sincerely say 'I am here now' and 'I am not Saul Kripke'.) However, although each of us bears the same relation to these primary intensions, Harold doesn't know what Gideon and I do, and Gideon's knowledge, unlike mine in the case of (13a), is not a priori. The other likely option is that what is known, and reported to be known, in these cases are the ordinary propositions which are the secondary intensions of (13a) and (13b) in my contexts. It might be maintained that Gideon and I know these propositions in virtue of accepting, and being justified in accepting, different meanings (characters) which assign our respective contexts the same propositions; whereas Harold fails to know these propositions because there is no meaning that he accepts (and is justified in accepting) which assigns these propositions to a context with him as agent. On this view, the reason my knowledge reported in (14a) arguably counts as a priori is that the meaning I accept requires no empirical justification for accepting; the reason Gideon's (reported in (15a)) doesn't is that his does require such justification.

Although this is not unreasonable, it contradicts rather than vindicates strong two-dimensionalism. In the case of the (a) sentences, the thing known by Gideon and me (but not Harold) is a single proposition that is both contingent and known a priori by me; thus there are propositions that are both contingent and knowable a priori. In the case of the (b) sentences, the thing known by Gideon and me is a single proposition that is both necessary and knowable only a posteriori. Moreover, since Harold fails to know it despite its necessity, propositions cannot be sets of possible world-states after all.

What, then, is to become of two-dimensionalism, as a semantic framework within which to attempt a descriptivist revival? As I see it, the most plausible option is to modify it by (i) identifying secondary intensions of sentences with structured propositions, and meanings of sentences with functions from contexts to structured propositions, (ii) recasting 'primary intensions' of sentences as meanings (i.e. characters) rather than propositions determined by them, (iii) analysing the relevant attitude ascriptions x v's that S-as ambiguous between (a) a standard reading in which the ascription is true relative to a context C iff the agent accepts some meaning M which assigns to the agent's context the structured proposition p which is the proposition semantically expressed by S relative to C, and (b) a de se reading in which it is true relative to C iff the agent accepts the meaning of S (see Lewis 1979; Perry 1977). According to this, weakened version of two-dimensionalism, a proposition p is both necessary and knowable only a posteriori iff (i) p is necessary, (ii) p is knowable in virtue of one's justifiably accepting some meaning M (and knowing that it expresses a truth), where M is such that (a) it assigns p to one's context, (b) it assigns a false proposition to some other context, and (c) one's justification for accepting M (and believing it to express a truth) requires one to

possess empirical evidence, and (iii) p is knowable only in this way. Similarly, p is both contingently true and knowable a priori iff in addition to being contingently true, p is knowable in virtue of one's justifiably accepting some meaning M (and knowing that it expresses a truth), where M is such that (*a*) it assigns p to one's context, (*b*) it assigns a true proposition to every context, and (*c*) one's justification for accepting M (and believing it to express a truth) does not require one to possess empirical evidence.

This system—*weak two-dimensionalism*—gives a two-dimensionalist account of (7) and (10), without being falsified by our arguments against strong twodimensionalism. But can it be used to vindicate descriptivism, and avoid Kripke's arguments about names and natural kind terms? To be successful, descriptivists must find reference-fixing descriptions that escape his semantic arguments; they must rigidify those descriptions (or give them wide scope) in a way that avoids his modal and epistemological arguments; and they must show that this can be done without sacrificing the (alleged) virtues of standard descriptivist solutions to the puzzles of Frege and Russell.

3.2 Why One Should Not Expect to Find Reference-Fixing Descriptions

Recall Jackson's a priori defence of descriptivism. He argued, in effect, that the claim that the referents of expressions are semantically fixed by descriptions is irrefutable, since any refutation requires clear intuitions about what refers to what in different situations, and these can only be explained as arising from reference-fixing descriptions semantically associated with expressions by speakers. I disagree. First, there are clear cases in which we have no trouble identifying the referent of a term t, even though it is clear that there is no reference-fixing description associated with t by speakers. Kaplan's example of the identical twins Castor and Pollux, discussed earlier, is a case in point. We have no trouble identifying Castor as the referent of his use of I, and Pollux as the referent of his, just as we have no trouble recognizing ourselves as referents of our own uses. This is so despite the fact that the referent of I is not semantically fixed, for any of us, by descriptions we semantically associate with it. If this is true of I, it is surely also true of now, and may be true of other expressions as well. Secondly, even in cases in which there may be descriptions picking out the referent of a term that are, in some sense, associated with it by speakers, it remains to be shown these descriptions play any role in its semantics. One can describe possible scenarios in which our intuitions tell us that speakers use the word *and* to mean disjunction, the material conditional, the property of being a necessary truth, or the property of being a philosopher. Even if one were to grant the assumption that these intuitions arise from some internalized theory T that

unconsciously guides us, it would not follow that the meaning of *and*—it's character in Kaplan's sense—is one that yields as content in a context whatever satisfies the relevant description extractable from T. Surely not every word is a descriptive indexical in Kaplan's sense. To miss this point is to miss the distinction between (i) semantic facts about what an expression means, or what its referent or content is in a context, and (ii) pre-semantic facts in virtue of which the expression has the meaning, and hence the referents and contents in different contexts, that it does. Whereas the descriptivist needs reference-fixing descriptions to be involved in (i), Jackson's argument can't exclude the possibility that their only role is in (ii). Finally, the claim that our ability to categorize cases in certain ways presupposes the sort of underlying knowledge required by the description theory is tendentious in something like the way that Plato's attribution of a priori knowledge of mathematics to the slave boy in the *Meno* is tendentious. There are other ways to explain the recognition of new facts.

For these reasons, it is an error to assume that descriptions semantically fixing the referents of names and natural kind terms must be available. Instead of looking for some a priori guarantee, one must consider candidate descriptions case by case. When one does this, the results are not promising. An often noted fact about names is the enormous variability in the descriptive information associated with the same name by different speakers. Although most speakers who have enough familiarity with a name to be able to use it possess some descriptive information about its referent, little, if any, of this information is common to all of them—certainly not enough to identify the referent uniquely. What is more, many speakers would not be able to articulate any uniquely identifying description. The same point applies to natural kind terms like *water*, for which Jackson suggests the reference-fixing description 'something like: belonging to the kind which most of the clear, potable samples, acquaintance with which lead [led?] to the introduction of the word "water" in our language [belong to]' (1998b: 212) This clearly won't do. First, in order to understand water, an ordinary speaker doesn't have to have a view about what led to its introduction into our language. Secondly, one doesn't have to know that samples of water are standardly clear and potable. Imagine an unusually unfortunate English speaker, brought up in dismal and restricted circumstances, who never drank water, never imagined that anyone else did, and whose only acquaintance with it was with a cloudy stream of water spilling out of a drainpipe from a laundry. This person might correctly use water to refer to water, and might say and know, just as we do, that water is used for washing, but not know that water is often clear and potable. Since such a speaker may well understand the word, and use it to designate instances of the same kind that we do, without associating it with Jackson's proposed description, that description is not part of its meaning, and does not qualify as semantically fixing its referent.

As indicated earlier, examples like these have led several descriptivists to embrace causal descriptivism, according to which the reference-fixing description for a name (or kind term) n is something like the thing I have heard of under the name "n" or, perhaps, the causal source of this token of "n", where, David Lewis reminds us, to find 'an account of the relation being invoked here, just consult the writings of causal theorists of reference'. There are three problems with this view. First, I might use Zaza to refer to a certain dog in the neighbourhood, having forgotten that I introduced the name myself, and wrongly thinking that I picked it up from someone else. Since in such a case I use the name to refer to the dog, though I may never have heard it used by anyone else, there is some difficulty with Lewis's first description. The second problem is common to both descriptions, and to certain versions of the casual-historical theory of reference from which they are extracted. As Jonathan McKeown-Green (2002) has pointed out, not all cases in which a speaker successfully uses a name n to refer are cases in which he has either introduced n himself, or acquired n from someone else with the intention of preserving the reference of his source. Suppose, for example, that one knows of a certain region in Ireland in which the residents of different towns see to it that there is always exactly one person bearing the name Patrick O'Grady. Learning of this curious fact, I set out to visit the region to interview the different men bearing that name. On entering a pub in a new town, I announce 'I am looking for Patrick O'Grady, whom I am willing to pay for an interview for my new book'. In so doing I successfully use the name to refer to the man, and say something about him-not because I have acquired the name through a causal-historical chain of reference transmission, but because I am able to speak the language of the community in which the referent of the name has already been established.

This brings us to the third, and most fundamental, problem with the attempt to appropriate causal-historical theories of reference transmission for the purposes of descriptivism. Egocentric, metalinguistic descriptions associated with names are no more parts of their meanings than similar egocentric, metalinguistic descriptions are parts of the meanings of other words in the language. Standardly, when a speaker uses any word-magenta, abode, osteopath, alphabetize, necessarily, etc.he intends to use it in accord with the linguistic conventions of the community. He intends to use it to refer to, or express, whatever other competent members of the community do. In the case of proper names, it is recognized both that a given name may be used by only a subpart of the community, and that different members of the relevant subcommunity (who use the name to refer to the same individual) may associate it with very different descriptive information without the name meaning something different for each of them. Thus, the general intention that one's use of words conform with the linguistic conventions of one's community translates, in the case of most names, into the intention to use them to refer to whomever or whatever other relevant members of the community use them to

refer to. Some such intention is a standard condition on normal language use, not a part of meaning.

There is a larger lesson here involving historical chains of reference transmission. On the picture one often gets, a name isn't part of my language at all until I either introduce it myself, or encounter someone else using it, and form the reference-fixing intention that in my idiolect it will refer to whatever it refers to in the idiolect of those from whom I acquire it. This picture is misleading. The language I speak is a common language, of which the name is part before I ever encounter it. As a speaker, I need not know all the linguistic properties of the words in my language; my knowledge is partial, just as my knowledge of other social institutions of which I am a part is partial. Nevertheless, since I am a competent member of the linguistic community, I can appropriate a word that may be new to me, and use it with the meaning and reference it has already acquired. In the case of a name, the word probably entered the language via the stipulation of some authorities—say the parents of a newborn child. It is retained in the language by a practice of various speakers using it to refer to that individual; and speakers normally encounter it for the first time by hearing others use it—everyone intending to use it with the meaning or reference it has already attained. If this picture is right, then historical chains, though they exist, are not themselves referencedetermining mechanisms (this conception of the proper way to view historical chains of reference transmission is developed in McKeown-Green 2002, ch. 9). Thus, no metalinguistic descriptions invoking them are going to play the role of semantically fixing the reference of names (or natural kind terms).

3.3 Why the Modal and Epistemological Arguments Can't Be Avoided

In this section we put aside problems about finding reference-fixing descriptions in order to focus on the problems that would confront the descriptivist, even if such descriptions could be found. In order to avoid Kripke's modal argument, such descriptions must either be rigidified or stipulated to take wide scope over modal operators. There are overwhelming difficulties with each of these options.

First, consider *actually* rigidified descriptions. Suppose (i) that *Saul Kripke* \neq *David Kaplan* is an example of the necessary a posteriori, (ii) that *the x: SKx* and *the x: DKx* are descriptions that semantically fix the referents of the two names, (iii) that *the x: SKx* \neq *the x: DKx* is contingent because there are world-states in which the two descriptions denote the same person, and (iv) that the contents of the two names are given by *the x: actually SKx* and *the x: actually DKx*. On these assumptions *Saul Kripke* \neq *David Kaplan* conforms to the theses of weak two-dimensionalism. However, the analysis is incorrect, because (iv) is incorrect. The problem with it is

REFERENCE AND DESCRIPTION 421

based on the fact that actual believers share many beliefs with merely possible believers. I, along with others, believe that Saul Kripke \neq David Kaplan; and it is not unreasonable to suppose that we also believe of the actual world-state @ that it is a world-state with respect to which it is true that Saul Kripke \neq David Kaplan. A similar point holds for merely possible believers. In some possible world-states w, various agents believe that Saul Kripke \neq David Kaplan; in addition, they believe of w that it is a world-state with respect to which it is true that Saul Kripke \neq David Kaplan. However, they need not have any beliefs about @. Supposing they don't, I would be wrong if I were now to say, *In w, they believe that the x: actually SKx* \neq *the x: actually Dkx*. Thus, if (iv) is correct, I must have been wrong in saying *In w, they believe that Saul Kripke* \neq David Kaplan. Since I wasn't wrong; (iv) isn't correct (this argument is presented and defended in detail in Soames 2002: 39–49).

When spelled out in detail, this argument makes use of the weak twodimensionalist assumption that, on a standard reading of *x* believes that *S*, the ascription is true with respect to a context C and world-state w iff in w the agent believes the proposition expressed by S in C—an assumption traditional descriptivists often employ when using Frege's puzzle and Russell's problem of negative existentials to 'refute' non-descriptive analyses. What the argument shows is that if the content of a name n is given by an *actually*-rigidified description, then, on this reading, a belief ascription containing n in the complement clause is true of an agent x with respect to a merely possible world-state w only if in w x believes certain things about (not w) but the world C_s of the context used to report x's belief. Since there is no such reading of English belief ascriptions, the descriptivist proponent of weak two-dimensionalism cannot take the contents of names to be given by *actually*rigidified descriptions.

The problem gets worse when one realizes (i) that the only plausible referencefixing descriptions to which the actuality operator might be attached contain indexicals referring to the speaker and/or his utterance and time of utterance³ and (ii) that the only remotely plausible candidates for such descriptions are variants of those put forward by Lewis and other causal descriptivists. For example, consider Lewis's *thing I have heard of under the name "Venus"* or *causal source of this token of "Venus"*. Under the reading of belief ascriptions just indicated, my utterance of

(17) The ancient Babylonians believed that Venus was a star

is true only if the ancients had views about me and which things I have heard of under which names, or about the causal sources of specific utterances of mine. Obviously, this is absurd; these ascriptions have no such readings. Nor do they have

³ Both me and my Twin Earth duplicate associate the same purely qualitative descriptions with n, while using it to refer to different things. If this is to be accounted for by reference-fixing descriptions, they will have to contain indexicals referring to particular contextual parameters.

the second, de se reading that the two-dimensionalist sometimes alleges them to have—namely, one in which (17) is true only if the ancient Babylonians accepted the character of the complement of (17), which, on the Lewis causal-descriptivist analysis, they would do only if they took themselves to have heard of some object under the name "Venus". In fact, my utterance of (17) is true, even though they were not familiar with the name "Venus", and so would not have accepted this character. Finally, the absurdity of combining this analysis of names with the de se reading of belief ascriptions posited by some two-dimensionalist is brought out by (18).

- (18*a*) My friends in Mexico City believe Henry has been badmouthing me.
- (18*b*) My friends in Mexico City believe that the x:actually I have heard of x under the name "Henry" has been badmouthing me.

On this analysis, (18a) is analysed as (18b), and hence is predicted to be true iff each friend in Mexico City accepts the character of the complement sentence, and so believes that someone he has heard of under the name "Henry" has been badmouthing him. This cannot be.

The lesson to be drawn is that even if the causal descriptivist could provide reference-fixing descriptions for names and natural kind terms, the semantic contents of these terms could not be given by rigidifying these descriptions using the actuality operator. What about using *dthat*? Although this would avoid some of the absurdities involving actually, others would remain, and one new problem would surface. The difficulties that remain concern the second, de se reading of attitude ascriptions posited by the two-dimensionalist. If the purportedly reference-fixing descriptions to which *dthat* is attached are—as they must be—egocentric, metalinguistic descriptions of the sort provided by causal descriptivist, then the absurdities just discussed involving ascriptions like (17) and (18) carry over to analyses employing *dthat* rather than *actually*. When one considers the standard reading in which attitude ascriptions report relations between agents and the propositions semantically expressed by their complement clauses, the situation changes. If names are taken to be *dthat*-rigidified descriptions, we get the desired result that it is possible to believe that Saul Kripke \neq David Kaplan without believing anything about the actual world-state, or other contextual parameters; but we also get the result that the semantic content of a name, relative to a context C, is just its referent in C, and so (i) co-referential names are substitutable without change in content or truthvalue in attitude ascriptions, and (ii) negative existentials involving so-called empty names are predicted either not to express propositions at all, or to express propositions with gaps in them. In short, this version of descriptivism faces Frege's puzzle and Russell's problem in essentially the same way that anti-descriptivist theories do. In effect, this version of descriptivism is equivalent to direct-reference theories of names to which one has added (a) dubious claims about descriptions semantically fixing the referents of names, and (b) the postulation of an extra reading of belief ascriptions containing names that they do not have. For these reasons, appealing to *dthat*-rigidified descriptions is problematic for the descriptivist.

The only remaining descriptivist alternative for dealing with the modal argument involves analysing names as (non-rigid) descriptions that are required to take wide scope over modal operators, while retaining small scope when they occur embedded under verbs of propositional attitude. To adopt this strategy is, in effect, to give up appealing to a two-dimensionalist framework, since the alleged difference between the behaviour of a sentence containing a name when embedded under a modal operator and its behaviour when embedded under an epistemic operator (verb of propositional attitude) is now attributed simply to scope. As before, the strategy can scarcely get off the ground because the only remotely plausible reference-fixing descriptions contain indexicals referring to the speaker and/or his utterance and utterance time; and the content of such a description is never what a name n contributes to the proposition expressed by *x* believes that *n* is *F*. Thus, this approach cannot plausibly account for elementary examples like (17). In addition, the strategy of assigning different scopes to the alleged descriptive contents of names embedded in modal and epistemic constructions leads to absurdities similar to those revealed by the arguments given above against strong two-dimensionalism (Soames 2002: 25-39).

4. PROSPECTS OF NON-DESCRIPTIVISM

Having raised difficulties for descriptivism, I close by pointing the reader to leading attempts to address the concerns that have motivated descriptivism, without embracing descriptive semantic analyses. The first concern involves the necessary a posteriori and the contingent a priori. Although some philosophers have found it difficult to understand how a single proposition can be both necessary and knowable only a posteriori, Kripke provided the needed clue at the very outset. Some properties—e.g. the property of being made of molecules, the property of being a human being, and the property of being not identical with me—are essential to anything that has them. We know this a priori about many properties, even though we can know of a particular that it has one of these properties only a posteriori. If P expresses such a property and o is an object that has it, then the proposition expressed by *If x exists, then x is P* relative to an assignment of o to "x" will be necessary, but knowable only a posteriori. Although there is no real mystery here, working this view out in appropriate generality and using it to put to rest objections to the necessary a posteriori is a subtle and sometimes tricky matter.

However, it can be done (see Soames 2005*b*, forthcoming). The connection between the contingent a priori and non-descriptivist semantics is more troublesome. Although there are genuine examples, like (10), that are both contingent and knowable a priori, it is (contra Kripke) doubtful that any such examples involve proper names or natural kind terms (see Salmon 1987–88; also Soames 2003*b*, ch. 16; 2005*b*).

The final set of concerns facing non-descriptive analyses of names are those posed by the puzzles of Frege and Russell. Among the semantically non-descriptive strategies for dealing with Frege's puzzle are those developed by Nathan Salmon (1986), David Braun (1998, 2002), Mark Richard (1990), Richard Larson and Peter Ludlow (1993), Mark Crimmins (1992), and me (2002, 2005*a*). In addition, Salmon has made substantial progress on Russell's problem of negative existentials (1987, 1998). So, although the traditional problems for directly referential accounts of names have not gone away, progress on them continues to be made.⁴

REFERENCES

Braun, David (1998). 'Understanding Belief Reports'. Philosophical Review, 107: 555-95.

---- (2002). 'Cognitive Significance, Attitude Ascriptions, and Ways of Believing Propositions'. *Philosophical Studies*, 108: 65–81.

Chalmers, David (1996). The Conscious Mind. New York: Oxford University Press.

(2000). 'On Sense and Intension', in J. E. Tomberlin (ed.), *Philosophical Perspectives*, 16: *Language and Mind*. Oxford: Blackwell.

Crimmins, Mark (1992). Talk About Beliefs. Cambridge, Mass.: MIT Press.

Davies, Martin, and Humberstone, Lloyd (1980). 'Two Notions of Necessity'. *Philosophical Studies*, 38: 1–30.

Dummett, Michael (1973). *Frege: Philosophy of Language*. New York: Harper & Row. Evans, Gareth (1979). 'Reference and Contingency'. *The Monist*, 62: 161–89.

Frege, Gottlob (1970). 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. Peter Geach and Max Black. Oxford: Blackwell.

Jackson, Frank (1998a). From Metaphysics to Ethics. Oxford: Oxford University Press.

—— (1998b). 'Reference and Description Revisited', in J. E. Tomberlin (ed.), Philosophical Perspectives, 12:Language, Mind and Ontology. Oxford: Blackwell.

Kaplan, David (1979). 'On the Logic of Demonstratives'. *Journal of Philosophical Logic*, 8: 81–98.

——(1989). 'Demonstratives', in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press.

Kripke, Saul (1979). 'A Puzzle About Belief'. Repr. in Peter Ludlow (ed.), *Readings in the Philosophy of Language*. Cambridge, Mass.: MIT Press, 1997.

—(1980). Naming and Necessity. Cambridge, Mass.: Harvard University Press.

⁴ Thanks to Jim Pryor and Ali Kazmi for helpful comments on an earlier draft. For a more thorough treatment of many of the issues discussed in this chapter, see Soames (2005*b*).

| Kroon, Fred (1987). 'Causal Descriptivism'. Australasian Journal of Philosophy, 65: 1–17. |
|---|
| Larson, Richard, and Ludlow, Peter (1993). 'Interpreted Logical Forms'. Synthese, 95 305–56. |
| Lewis, David (1979). 'Attitudes De Dicto and De Se'. Repr. in Lewis, Philosophical Papers |
| New York: Oxford University Press, 1983. |
| —— (1984). 'Putnam's Paradox'. Repr. in Lewis, <i>Papers in Metaphysics and Epistemology</i> Cambridge: Cambridge University Press, 1999. |
| ——(1996). 'Elusive Knowledge'. Repr. in Lewis, <i>Papers in Metaphysics and Epistemology</i> Cambridge: Cambridge University Press, 1999. |
| — (1997). 'Naming the Colours'. Repr. in Lewis, <i>Papers in Metaphysics and Epistemology</i> Cambridge: Cambridge University Press, 1999. |
| Ludlow, Peter, and Larson, Richard (1990). 'Interpreted Logical Forms'. Synthese |
| 95: 305–56. |
| McKeown-Green, Jonathan (2002). The Primacy of Public Language. Ph. D. diss., Princetor University. |
| Perry, John (1977). 'Frege on Demonstratives'. Philosophical Review, 86: 474–97. |
| (1979). 'The Problem of the Essential Indexical'. Noûs, 13: 3-21. |
| Putnam, Hilary (1970). 'Is Semantics Possible?' Repr. in Putnam, Philosophical Papers, ii |
| Cambridge: Cambridge University Press. |
| —— (1973). 'Explanation and Reference'. Repr. in Putnam, <i>Philosophical Papers</i> , ii Cambridge: Cambridge University Press. |
| —— (1975). 'The Meaning of "Meaning"'. Repr. in Putnam, <i>Philosophical Papers</i> , ii Cambridge: Cambridge University Press. |
| Richard, Mark (1990). Propositional Attitudes. Cambridge: Cambridge University Press. |
| Russell, Bertrand (1905). 'On Denoting'. Mind, 14: 479-493. |
| Salmon, Nathan (1986). Frege's Puzzle. Cambridge, Mass.: MIT Press. |
| —— (1987). 'Existence', in J. E. Tomberlin (ed.), <i>Philosophical Perspectives</i> , 1: <i>Metaphysics</i> Atascadero, Calif.: Ridgeview. |
| (1987-8). 'How to Measure the Standard Metre'. Proceedings of the Aristotelian Society |
| 88: 193–217. |
| ——(1998). 'Nonexistence'. <i>Noûs</i> , 32: 277–319. |
| Searle, John (1983). Intentionality. Cambridge: Cambridge University Press. |
| Soames, Scott (2002). Beyond Rigidity. New York: Oxford University Press. |
| (2003a). Philosophical Analysis in the Twentieth Century, i. Princeton: Princeton |
| University Press. |
| (2003b). Philosophical Analysis in the Twentieth Century, ii. Princeton: Princeton |
| University Press. |
| —— (2004). 'Knowledge of Manifest Kinds'. Facta Philosophica, 6: 159–81. |
| (2005a). 'Naming and Assertin', in Z. Szabó(ed.), Semantics vs. Pragmatics. Oxford |
| Oxford Universivty Press. |

15-Jackson-Chap-15.qxd

17/5/05

5:59 PM

Page 42

- ——(2005*b*). *Reference and Description: The Case against Two-Dimensionalism*. Princeton: Princeton University Press.
- (forth coming). 'Kripke, the Necessary Aposteriori, and the Two-Dimensionalist Heresy', in M. Garcia-Carpintero and J. Macia (eds.), *The Two-Dimensionalist Framework: Foundations and Applications*. Oxford: Oxford University Press.
- —— (forthcoming *a*). 'Naming and Asserting', in Z. Szabó (ed.), *Semantics vs. Pragmatics*. Oxford: Oxford University Press.

Soames, Scott (forthcoming *b*). *Reference and Description: The Two-Dimensionalist Attempt to Revive Descriptivism.* Princeton: Princeton University Press.

—— (in preparation). 'Knowledge of Manifest Natural Kinds'.

Stalnaker, Robert (1978). 'Assertion'. Repr. in Stalnaker, *Context and Content*. Oxford: Oxford University Press, 1999.

(1984). Inquiry. Cambridge, Mass.: MIT Press.