

## Demonstratives

# An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals

David Kaplan<sup>1</sup>

---

<sup>1</sup>This paper was prepared for and read (with omissions) at a symposium on Demonstratives at the March 1977 meetings of the Pacific Division of the American Philosophical Association. The commentators were Paul Benacerraf and Charles Chastain. Much of the material, including the formal system of section XVIII, was originally presented in a series of lectures at the fabled 1971 Summer Institute in the Philosophy of Language held at the University of California, Irvine. © 1977 by David Kaplan.

## Table of Contents

	Preface	483
I.	Introduction	489
II.	Demonstratives, Indexicals, and Pure Indexicals	489
III.	Two Obvious Principles	492
IV.	Remarks on Rigid Designators	492
V.	Argument for Principle 2: Pure Indexicals	498
VI.	Terminological Remarks	500
VI. (i)	Content and Circumstance	500
VI. (ii)	Character	505
VII.	Earlier Attempts: Index Theory	507
VIII.	Monsters Begat by Elegance	510
IX.	Argument for Principle 2: True Demonstratives	512
IX. (i)	The Arguments	513
IX. (ii)	The Fregean Theory of Demonstrations	514
IX. (iii)	The Fregean Theory of Demonstratives	516
IX. (iv)	Argument Against the Fregean Theory of Demonstratives	516
X.	Fixing the Reference vs. Supplying a Synonym	518
X. (i)	Reichenbach on Token Reflexives	519
XI.	The Meaning of Indexicals	520
XII.	Dthat	521
XIII.	Contexts, Truth, and Logical Truth	522
XIV.	Summary of Findings (so far): Pure Indexicals	523
XV.	Further Details: Demonstratives and Demonstrations	524
XVI.	Alternative Treatments of Demonstrations	527
XVII.	Epistemological Remarks	529
XVIII.	The Formal System	541
XIX.	Remarks on the Formal System	546
XX.	Adding 'Says'	553
XXI.	Russell on Egocentric Particulars and Their Dispensability	557
XXII.	On Proper Names	558

## Preface

In about 1966 I wrote a paper about quantification into epistemological contexts. There are very difficult metaphysical, logical, and epistemological problems involved in providing a treatment of such idioms which does not distort our intuitions about their proper use and which is up to contemporary logical standards. I did not then, and do not now, regard the treatment I provided as fully adequate. And I became more and more intrigued with problems centering on what I would like to call the *semantics of direct reference*. By this I mean theories of meaning according to which certain singular terms refer directly without the mediation of a Fregean *Sinn* as meaning. If there are such terms, then the proposition expressed by a sentence containing such a term would involve individuals directly rather than by way of the "individual concepts" or "manners of presentation" I had been taught to expect. Let us call such putative singular terms (if there are any) *directly referential terms* and such putative propositions (if there are any) *singular propositions*. Even if English contained no singular terms whose proper semantics was one of direct reference, could we determine to introduce such terms? And even if we had no directly referential terms and introduced none, is there a need or use for singular propositions?

The feverish development of quantified modal logics, more generally, of quantified intensional logics, of the 1960s gave rise to a metaphysical and epistemological malaise regarding the problem of identifying individuals across worlds—what, in 1967, I called the problem of "Trans-World Heir Lines." This problem was really just the problem of singular propositions: those which involve individuals directly, rearing its irrepressible head in the possible-world semantics that were then (and are now) so popular.

It was not that according to those semantical theories any sentences of the languages being studied were themselves taken to express singular propositions, it was just that singular propositions seemed to be needed in the analysis of the nonsingular propositions expressed by these sentences. For example, consider

$$(0) \quad \exists x(Fx \wedge \sim \Box Fx).$$

This sentence would not be taken by anyone to express a singular proposition. But in order to evaluate the truth-value of the component

$$\Box Fx$$

(under some assignment of an individual to the variable 'x'), we must first determine whether the *proposition* expressed by its component

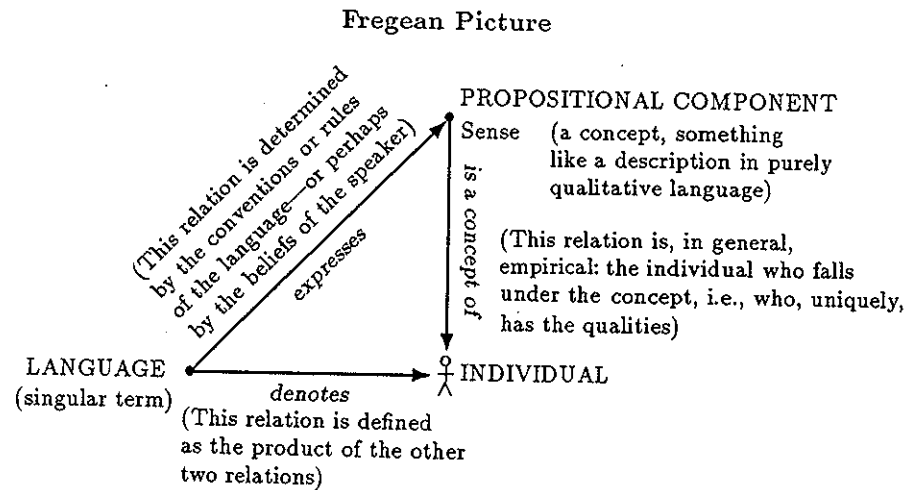
$Fx$

(under an assignment of an individual to the variable 'x') is a necessary proposition. So in the course of analyzing (0), we are required to determine the proposition associated with a formula containing a *free* variable. Now free variables under an assignment of values are paradigms of what I have been calling *directly referential* terms. In determining a semantical value for a formula containing a free variable we may be given a *value* for the variable—that is, an individual drawn from the universe over which the variable is taken to range—but nothing more. A variable's first and only meaning is its value. Therefore, if we are to associate a *proposition* (not merely a truth-value) with a formula containing a free variable (with respect to an assignment of a value to the variable), that proposition seems bound to be singular (even if valiant attempts are made to disguise this fact by using constant functions to imitate individual concepts). The point is, that if the component of the proposition (or the step in the construction of the proposition) which corresponds to the singular term is determined by the individual and the individual is directly determined by the singular term—rather than the individual being determined by the component of the proposition, which is directly determined by the singular term—then we have what I call a singular proposition. [Russell's semantics was like the semantical theories for quantified intensional logics that I have described in that although no (closed) sentence of *Principia Mathematica* was taken to stand for a singular proposition, singular propositions are the essential building blocks of all propositions.]

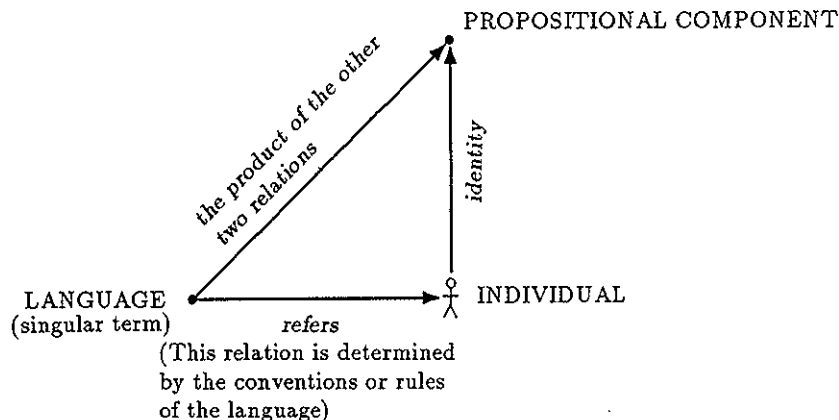
The most important hold-out against semantical theories that required singular propositions is Alonzo Church, the great modern champion of Frege's semantical theories. Church also advocates a version of quantified intensional logic, but with a subtle difference that finesses the need for singular propositions. (In Church's logic, given a sentential formula containing free variables and given an assignment of values to the variables, no proposition is yet determined. An additional assignment of "senses" to the free variables must be made before a proposition can be associated with the formula.) It is no accident that Church rejects *direct reference* semantical theories. For if there were singular terms which referred directly, it seems likely that Frege's problem: how can ' $\alpha = \beta$ ', if true, differ in meaning from ' $\alpha = \alpha$ ', could be reinstated,

while Frege's solution: that  $\alpha$  and  $\beta$ , though referring to the same thing, do so by way of different senses, would be blocked. Also: because of the fact that the component of the proposition is being determined by the individual rather than vice versa, we have something like a violation of the famous Fregean dictum that *there is no road back* from denotation to sense [propositional component]. (Recently, I have come to think that if we countenance singular propositions, a collapse of Frege's intensional ontology into Russell's takes place.)

I can draw some little pictures to give you an idea of the two kinds of semantical theories I want to contrast.



## Direct Reference Picture



(These pictures are not entirely accurate for several reasons, among them; that the contrasting pictures are meant to account for more than just singular terms and that the relation marked 'refers' may already involve a kind of Fregean sense used to fix the referent.)

I won't go into the pros and cons of these two views at this time. Suffice it to say that I had been raised on Fregean semantics and was sufficiently devout to wonder whether the kind of quantification into modal and epistemic contexts that seemed to require singular propositions really made sense. (My paper "Quantifying In" can be regarded as an attempt to *explain away* such idioms for epistemic contexts.)<sup>2</sup>

But there were pressures from quarters other than quantified intensional logic in favor of a semantics of direct reference. First of all there was Donnellan's fascinating paper "Reference and Definite Descriptions."<sup>3</sup> Then there were discussions I had had with Putnam in 1968 in which he argued with respect to certain natural kind terms like 'tiger' and 'gold', that if their Fregean senses were the kind of thing that one grasped when one understood the terms, then such senses could

<sup>2</sup>David Kaplan, "Quantifying In," *Synthese* 19 (1968): 178–214; reprinted in *The Philosophy of Language*, ed. A. P. Martinich (Oxford: Oxford University Press, 1985).

<sup>3</sup>Keith Donnellan, "Reference and Definite Descriptions," *Philosophical Review* 75 (1966): 281–304; reprinted in Martinich, *op. cit.*

not determine the extension of the terms. And finally Kripke's Princeton lectures of spring 1970, later published as *Naming and Necessity*,<sup>4</sup> were just beginning to leak out along with their strong attack on the Fregean theory of proper names and their support of a theory of direct reference.

As I said earlier, I was intrigued by the semantics of direct reference, so when I had a sabbatical leave for the year 1970–71, I decided to work in the area in which such a theory seemed most plausible: demonstratives. In fall 1970, I wrote, for a conference at Stanford, a paper "Dthat."<sup>5</sup> Using Donnellan's ideas as a starting point, I tried to develop the contrast between Fregean semantics and the semantics of direct reference, and to argue that demonstratives—although they *could* be treated on a Fregean model—were more interestingly treated on a direct reference model. Ultimately I came to the conclusion that something analogous to Donnellan's referential use of a definite description could be developed using my new demonstrative, "dthat." In the course of this paper I groped my way to a formal semantics for demonstratives rather different in conception from those that had been offered before.

In spring 1971, I gave a series of lectures at Princeton on the semantics of direct reference. By this time I had seen a transcript of *Naming and Necessity* and I tried to relate some of my ideas to Kripke's.<sup>6</sup> I also had written out the formal semantics for my Logic of Demonstratives. That summer at the Irvine Philosophy of Language Institute I lectured again on the semantics of direct reference and repeated some of these lectures at various institutions in fall 1971. And there the matter has stood except for a bit of updating of the 1971 Logic of Demonstratives notes in 1973.

I now think that demonstratives can be treated correctly only on a direct reference model, but that my earlier lectures at Princeton and Irvine on direct reference semantics were too broad in scope, and that the most important and certainly the most convincing part of my theory is just the logic of demonstratives itself. It is based on just a few quite

<sup>4</sup>Saul Kripke, "Naming and Necessity," in *Semantics of Natural Language*, ed. G. Harman and D. Davidson (Dordrecht: Reidel, 1972); revised edition published as a separate monograph, *Naming and Necessity* (Oxford: Basil Blackwell, 1980). References are to the revised edition.

<sup>5</sup>David Kaplan, "Dthat," in *Syntax and Semantics*, vol. 9, ed. P. Cole (New York: Academic Press, 1978); reprinted in Martinich, *op. cit.*

<sup>6</sup>Although the central ideas of my theory had been worked out before I became familiar with *Naming and Necessity*, I have enthusiastically adopted the 'analytical apparatus' and some of the terminology of that brilliant work.

simple ideas, but the conceptual apparatus turns out to be surprisingly rich and interesting. At least I hope that you will find it so.

In this work I have concentrated on pedagogy. Philosophically, there is little here that goes beyond the Summer Institute Lectures, but I have tried, by limiting the scope, to present the ideas in a more compelling way. Some new material appears in the two speculative sections: XVII (Epistemological Remarks) and XX (Adding 'Says'). It is my hope that a theory of demonstratives will give us the tools to go on in a more sure-footed way to explore the *de.re* propositional attitudes as well as other semantical issues.

## I. Introduction

I believe my theory of demonstratives to be uncontroversial and largely uncontroversial. This is not a tribute to the power of my theory but a concession of its obviousness. In the past, no one seems to have followed these obvious facts out to their obvious consequences. I do that. What is original with me is some terminology to help fix ideas when things get complicated. It has been fascinating to see how interesting the obvious consequences of obvious principles can be.<sup>7</sup>

## II. Demonstratives, Indexicals, and Pure Indexicals

I tend to describe my theory as 'a theory of demonstratives', but that is poor usage. It stems from the fact that I began my investigations by asking what is said when a speaker points at someone and says, "He is suspicious."<sup>8</sup> The word 'he', so used, is a demonstrative, and the accompanying pointing is the requisite associated demonstration. I hypothesized a certain semantical theory for such demonstratives, and then I invented a new demonstrative, 'dthat', and stipulated that its semantics be in accord with my theory. I was so delighted with this methodological sleight of hand for my demonstrative 'dthat', that when I generalized the theory to apply to words like 'I', 'now', 'here', etc.—words which do *not* require an associated demonstration—I continued to call my theory a 'theory of demonstratives' and I referred to these words as 'demonstratives'.

That terminological practice conflicts with what I preach, and I will try to correct it. (But I tend to backslide.)

The group of words for which I propose a semantical theory includes the pronouns 'I', 'my', 'you', 'he', 'his', 'she', 'it', the demonstrative pronouns 'that', 'this', the adverbs 'here', 'now', 'tomorrow', 'yesterday', the adjectives 'actual', 'present', and others. These words have uses other than those in which I am interested (or, perhaps, depending on how you individuate words, we should say that they have homonyms in which I am not interested). For example, the pronouns 'he' and 'his' are used not as demonstratives but as bound variables in

<sup>7</sup>Not everything I assert is part of my theory. At places I make judgments about the correct use of certain words and I propose detailed analyses of certain notions. I recognize that these matters may be controversial. I do not regard them as part of the basic, obvious, theory.

<sup>8</sup>See "Dthat," p. 320 in Martinich.

For what is a man profited, if he shall gain  
the whole world, and lose his own soul?

What is common to the words or usages in which I am interested is that the referent is dependent on the context of use and that the meaning of the word provides a rule which determines the referent in terms of certain aspects of the context. The term I now favor for these words is 'indexical'. Other authors have used other terms; Russell used 'egocentric particular' and Reichenbach used 'token reflexive'. I prefer 'indexical' (which, I believe, is due to Pierce) because it seems less theory laden than the others, and because I regard Russell's and Reichenbach's theories as defective.

Some of the indexicals require, in order to determine their referents, an associated demonstration: typically, though not invariably, a (visual) presentation of a local object discriminated by a pointing.<sup>9</sup> These indexicals are the true demonstratives, and 'that' is their paradigm. The demonstrative (an expression) refers to that which the demonstration demonstrates. I call that which is demonstrated the 'demonstratum'.

A demonstrative without an associated demonstration is incomplete. The linguistic rules which govern the use of the true demonstratives 'that', 'he', etc., are not sufficient to determine their referent in all contexts of use. Something else—an associated demonstration—must be provided. The linguistic rules assume that such a demonstration accompanies each (demonstrative) use of a demonstrative. An incomplete demonstrative is not *vacuous* like an improper definite description. A demonstrative *can* be vacuous in various cases. For example, when its associated demonstration has no demonstratum (a hallucination)—or the wrong kind of demonstratum (pointing to a flower and saying 'he' in the belief that one is pointing to a man disguised as a flower<sup>10</sup>)—or too many demonstrata (pointing to two intertwined vines and saying

<sup>9</sup>However, a demonstration may also be opportune and require no special action on the speaker's part, as when someone shouts "Stop that man" while only one man is rushing toward the door. My notion of a demonstration is a theoretical concept. I do not, in the present work, undertake a detailed 'operational' analysis of this notion although there are scattered remarks relevant to the issue. I do consider, in XVI below, some alternative theoretical treatments of demonstrations.

<sup>10</sup>I am aware (1) that in some languages the so-called masculine gender pronoun may be appropriate for flowers, but it is not so in English; (2) that a background story can be provided that will make pointing at the flower a contextually appropriate, though deviant, way of referring to a man; for example, if we are talking of great hybridizers; and (3) that it is possible to treat the example as a *referential use* of the demonstrative 'he' on the model of Donnellan's referential use of a definite description (see "Reference and Definite Descriptions"). Under the referential use

'that vine'). But it is clear that one can distinguish a demonstrative with a vacuous demonstration: no referent; from a demonstrative with no associated demonstration: incomplete.

All this is by way of contrasting true demonstratives with pure indexicals. For the latter, *no associated demonstration is required, and any demonstration supplied is either for emphasis or is irrelevant*.<sup>11</sup> Among the pure indexicals are 'I', 'now', 'here' (in one sense), 'tomorrow', and others. The linguistic rules which govern *their* use fully determine the referent for each context.<sup>12</sup> No supplementary actions or intentions are needed. The speaker refers to himself when he uses 'I', and no pointing to another or believing that he is another or intending to refer to another can defeat this reference.<sup>13</sup>

Michael Bennett has noted that some indexicals have both a pure and a demonstrative use. 'Here' is a pure indexical in

I am in here

and is a demonstrative in

In two weeks, I will be here [pointing at a city on a map].

---

treatment we would assign as referent for 'he' whatever the speaker *intended* to demonstrate. I intended the example to exemplify a failed demonstration, thus, a case in which the speaker, falsely believing the flower to be some man or other in disguise, but having no particular man in mind, and certainly not intending to refer to anything other than that man, says, pointing at the flower, "He has been following me around all day."

<sup>11</sup>I have in mind such cases as pointing at oneself while saying 'I' (emphasis) or pointing at someone else while saying 'I' (irrelevance or madness or what?).

<sup>12</sup>There are certain uses of pure indexicals that might be called 'messages recorded for later broadcast', which exhibit a special uncertainty as to the referent of 'here' and 'now'. If the message: "I am not here now" is recorded on a telephone answering device, it is to be assumed that the time referred to by 'now' is the time of playback rather than the time of recording. Donnellan has suggested that if there were typically a significant lag between our production of speech and its audition (for example, if sound traveled very very slowly), our language might contain two forms of 'now': one for the time of production, another for the time of audition. The indexicals 'here' and 'now' also suffer from vagueness regarding the size of the spatial and temporal neighborhoods to which they refer. These facts do not seem to me to slur the difference between demonstratives and pure indexicals.

<sup>13</sup>Of course it is certain intentions on the part of the speaker that make a particular vocable the first person singular pronoun rather a nickname for Irving. My semantical theory is a theory of word meaning, not speaker's meaning. It is based on linguistic rules known, explicitly or implicitly, by all competent users of the language.

### III. Two Obvious Principles

So much for preliminaries. My theory is based on two obvious principles. The first has been noted in every discussion of the subject.

**Principle 1** *The referent of a pure indexical depends on the context, and the referent of a demonstrative depends on the associated demonstration.*

If you and I both say 'I' we refer to different persons. The demonstratives 'that' and 'he' can be correctly used to refer to any one of a wide variety of objects simply by adjusting the accompanying demonstration.

The second obvious principle has less often been formulated explicitly.

**Principle 2** *Indexicals, pure and demonstrative alike, are directly referential.*

### IV. Remarks on Rigid Designators

In an earlier draft I adopted the terminology of Kripke, called indexicals 'rigid designators', and tried to explain that my usage differed from his. I am now shying away from that terminology. But because it is so well known, I will make some comments on the notion or notions involved.

The term 'rigid designator' was coined by Saul Kripke to characterize those expressions which designate the same thing in every possible world in which that thing exists and which designate nothing elsewhere. He uses it in connection with his controversial, though, I believe, correct claim that proper names, as well as many common nouns, are rigid designators. There is an unfortunate confusion in the idea that a proper name would designate nothing if the bearer of the name were not to exist.<sup>14</sup> Kripke himself adopts positions which seem inconsistent with this feature of rigid designators. In arguing that the object designated by a rigid designator need not exist in every possible world, he seems to assert that under certain circumstances what is expressed by 'Hitler does not exist' would have been true, and not because 'Hitler' would have designated nothing (in *that* case we might have given the sentence *no* truth-value) but because what 'Hitler' would have designated—namely

Hitler—would not have existed.<sup>15</sup> Furthermore, it is a striking and important feature of the possible world semantics for quantified intensional logics, which Kripke did so much to create and popularize, that variables, those paradigms of rigid designation, designate the same individual in *all* possible worlds whether the individual "exists" or not.<sup>16</sup>

Whatever Kripke's intentions (did he, as I suspect, misdescribe his own concept?) and whatever associations or even meaning the phrase 'rigid designator' may have, I intend to use '*directly referential*' for an expression whose referent, once determined, is taken as fixed for all possible circumstances, i.e., is taken as *being* the propositional component.

For me, the intuitive idea is not that of an expression which *turns out* to designate the same object in all possible circumstances, but an expression whose semantical *rules* provide *directly* that the referent in all possible circumstances is fixed to be the actual referent. In typical cases the semantical rules will do this only implicitly, by providing a way of determining the *actual* referent and no way of determining any other propositional component.<sup>17</sup>

We should beware of a certain confusion in interpreting the phrase 'designates the same object in all circumstances'. We do not mean that the expression *could not have been used* to designate a different object.

<sup>15</sup> Kripke, *Naming and Necessity*, p. 78.

<sup>16</sup> The matter is even more complicated. There are two 'definitions' of 'rigid designator' in *Naming and Necessity*, pp. 48–49. The first conforms to what seems to me to have been the intended concept—same designation in *all* possible worlds—the second, scarcely a page later, conforms to the more widely held view that a rigid designator need not designate the object, or any object, at worlds in which the object does not exist. According to this conception a designator cannot, at a given world, designate something which does not exist in that world. The introduction of the notion of a *strongly* rigid designator—a rigid designator whose designatum exists in all possible worlds—suggests that the latter idea was uppermost in Kripke's mind. (The second definition is given, unequivocally, on page 146 of "Identity and Necessity," in *Identity and Individuation*, ed. M. K. Munitz (New York: New York University Press, 1971).) In spite of the textual evidence, systematic considerations, including the fact that variables cannot be accounted for otherwise, leave me with the conviction that the former notion was intended.

<sup>17</sup> Here, and in the preceding paragraph, in attempting to convey my notion of a directly referential singular term, I slide back and forth between two metaphysical pictures: that of possible worlds and that of structured propositions. It seems to me that a truly semantical idea should presuppose neither picture, and be expressible in terms of either. Kripke's discussion of rigid designators is, I believe, distorted by an excessive dependence on the possible worlds picture and the associated semantical style. For more on the relationship between the two pictures, see pages 724–25 of my "How to Russell a Frege-Church," *The Journal of Philosophy* 72 (1975): 716–29.

<sup>14</sup> I have discussed this and related issues in "Bob and Carol and Ted and Alice," in *Approaches to Natural Language*, ed. J. Hintikka et al. (Dordrecht: Reidel, 1973), especially appendix X.

We mean rather that given a *use* of the expression, we may ask of *what has been said* whether *it* would have been true or false in various counterfactual circumstances, and in such counterfactual circumstances, which are the individuals relevant to determining truth-value. Thus we must distinguish possible occasions of *use*—which I call *contexts*—from possible circumstances of *evaluation* of what was said on a given occasion of use. Possible circumstances of evaluation I call circumstances or, sometimes, just *counterfactual situations*. A directly referential term may designate different objects when used in different *contexts*. But when evaluating what was said in a given context, only a single object will be relevant to the evaluation in all circumstances. This sharp distinction between *contexts of use* and *circumstances of evaluation* must be kept in mind if we are to avoid a seeming conflict between Principles 1 and 2.<sup>18</sup> To look at the matter from another point of view, once we recognize the obviousness of both principles (I have not yet argued for Principle 2) the distinction between contexts of use and circumstances of evaluation is forced upon us.

If I may wax metaphysical in order to fix an image, let us think of the vehicles of evaluation—the what-is-said in a given context—as propositions: Don't think of propositions as sets of possible worlds, but rather as structured entities looking something like the sentences which express them. For each occurrence of a singular term in a sentence there will be a corresponding constituent in the proposition expressed. The constituent of the proposition determines, for each circumstance of evaluation, the object relevant to evaluating the proposition in that circumstance. In general, the constituent of the proposition will be some sort of complex, constructed from various attributes by logical composition. But in the case of a singular term which is directly referential, the constituent of the proposition is just the object itself. Thus it is that it does not just *turn out* that the constituent determines the same object in every circumstance, the constituent (corresponding to a rigid designator) just *is* the object. *There is no determining to do at all*. On this picture—and this is *really* a picture and not a theory—the definite description

- (1) The  $n[(\text{Snow is slight} \wedge n^2 = 9) \vee (\sim \text{Snow is slight} \wedge 2^2 = n + 1)]$ <sup>19</sup>

<sup>18</sup>I think it likely that it was just the failure to notice this distinction that led to a failure to recognize Principle 2. Some of the history and consequences of the conflation of Context and Circumstance is discussed in section VII.

<sup>19</sup>I would have used 'snow is white', but I wanted a contingent clause, and so many

would yield a constituent which is complex although it would determine the same object in all circumstances. Thus, (1), though a rigid designator, is not directly referential from this (metaphysical) point of view. Note, however, that every proposition which contains the complex expressed by (1) is *equivalent* to some singular proposition which contains just the number three itself as constituent.<sup>20</sup>

The semantical feature that I wish to highlight in calling an expression *directly referential* is not the *fact* that it designates the same object in every circumstance, but the *way* in which it designates an object in any circumstance. Such an expression is a *device of direct reference*. This does not imply that it has no conventionally fixed semantical rules which determine its referent in each context of use; quite the opposite. There are semantical rules which determine the referent in each context of use—but that is all. *The rules do not provide a complex which together with a circumstance of evaluation yields an object. They just provide an object.*

If we keep in mind our sharp distinction between contexts of use and circumstances of evaluation, we will not be tempted to confuse a rule which assigns an object to each *context* with a 'complex' which assigns an object to each *circumstance*. For example, each context has an *agent* (loosely, a speaker). Thus an appropriate designation rule for a directly referential term would be:

- (2) In each possible context of use the given term refers to the agent of the context.

But this rule could not be used to assign a relevant object to each circumstance of evaluation. Circumstances of evaluation do not, in general, have agents. Suppose I say,

- (3) I do not exist.

Under what circumstances would *what I said* be true? It would be true in circumstances in which I did not exist. Among such circumstances are those in which no one, and thus, no speakers, no agents exist. To search a circumstance of evaluation for a speaker in order to (mis)apply rule (2) would be to go off on an irrelevant chase.

people (possibly including me) nowadays seem to have views which allow that 'snow is white' may be necessary.

<sup>20</sup>I am ignoring propositions expressed by sentences containing epistemic operators or others for which equivalence is not a sufficient condition for interchange of operand.



Three paragraphs ago I sketched a metaphysical picture of the structure of a proposition. The picture is taken from the semantical parts of Russell's *Principles of Mathematics*.<sup>21</sup> Two years later, in "On Denoting,"<sup>22</sup> even Russell rejected that picture. But I still like it. It is not a part of my theory, but it well conveys my conception of a directly referential expression and of the semantics of direct reference. (The picture needs *some* modification in order to avoid difficulties which Russell later noted—though he attributed them to Frege's theory rather than his own earlier theory.)<sup>23</sup>

If we adopt a possible worlds semantics, all directly referential terms

<sup>21</sup>Bertrand Russell, *The Principles of Mathematics* (London: Allen & Unwin, 1903).

<sup>22</sup>Bertrand Russell, "On Denoting," *Mind* 14 (1905): 479–93.

<sup>23</sup>Here is a difficulty in Russell's 1903 picture that has some historical interest. Consider the proposition expressed by the sentence, 'The centre of mass of the Solar System is a point'. Call the proposition, 'P'. P has in its subject place a certain complex, expressed by the definite description. Call the complex, 'Plexy'. We can describe Plexy as "the complex expressed by 'the center of mass of the solar system'." Can we produce a directly referential term which designates Plexy? Leaving aside for the moment the controversial question of whether 'Plexy' is such a term, let us imagine, as Russell believed, that we can directly refer to Plexy by affixing a kind of *meaning marks* (on the analogy of quotation marks) to the description itself. Now consider the sentence "'the center of mass of the solar system' is a point'. Because the subject of this sentence is directly referential and refers to Plexy, the proposition the sentence expresses will have as its subject constituent Plexy itself. A moment's reflection will reveal that this proposition is simply P again. But this is absurd since the two sentences speak about radically different objects.

(I believe the foregoing argument lies behind some of the largely incomprehensible arguments mounted by Russell against Frege in "On Denoting," though there are certainly other difficulties in that argument. It is not surprising that Russell there confused Frege's theory with his own of *Principle of Mathematics*. The first footnote of "On Denoting" asserts that the two theories are "very nearly the same.")

The solution to the difficulty is simple. Regard the 'object' places of a singular proposition as marked by some operation which cannot mark a complex. (There always will be some such operation.) For example, suppose that no complex is (represented by) a set containing a single member. Then we need only add { . . . } to mark the places in a singular proposition which correspond to directly referential terms. We no longer need worry about confusing a complex with a propositional constituent corresponding to a directly referring term because no complex will have the form {x}. In particular, Plexy ≠ {Plexy}. This technique can also be used to resolve another confusion in Russell. He argued that a sentence containing a nondenoting directly referential term (he would have called it a nondenoting 'logically proper name') would be meaningless, presumably because the purported singular proposition would be incomplete. But the braces themselves can fill out the singular proposition, and if they contain nothing, no more anomalies need result than what the development of Free Logic has already inured us to.

will be regarded as rigid designators in the *modified* sense of an expression which designates the same thing in *all* possible worlds (irrespective of whether the thing exists in the possible world or not).<sup>24</sup> However, as already noted, I do not regard all rigid designators—not even all strongly rigid designators (those that designate something that exists in all possible worlds) or all rigid designators in the modified sense—as directly referential. I believe that proper names, like variables, are directly referential. They are not, in general, strongly rigid designators nor are they rigid designators in the original sense.<sup>25</sup> What is characteristic of directly referential terms is that the designatum (referent) determines the propositional component rather than the propositional component, along with a circumstance, determining the designatum. It is for this reason that a directly referential term that designates a contingently existing object will still be a rigid designator in the modified sense. The propositional component need not choose its designatum from those offered by a passing circumstance; it has already secured its designatum before the encounter with the circumstance.

When we think in terms of possible world semantics this fundamental distinction becomes subliminal. This is because the style of the semantical rules obscures the distinction and makes it appear that directly referential terms differ from ordinary definite descriptions only in that the propositional component in the former case must be a *constant* function of circumstances. In actual fact, the referent, in a circumstance, of a directly referential term is simply *independent* of the circumstance and is no more a function (constant or otherwise) of circumstance, than my action is a function of your desires when I decide to do it whether you like it or not. The distinction that is obscured by the style of possible world semantics is dramatized by the structured propositions picture. That is part of the reason why I like it.

Some directly referential terms, like proper names, may have no semantically relevant descriptive meaning, or at least none that is specific: that distinguishes one such term from another. Others, like the indexicals, may have a limited kind of specific descriptive meaning relevant to the features of a context of use. Still others, like 'dthat' terms (see below), may be associated with full-blown Fregean senses used to fix the referent. But in any case, the descriptive meaning of a directly referential term is no part of the propositional content.

<sup>24</sup>This is the *first sense* of footnote 16.

<sup>25</sup>This is the *second sense* of footnote 16.

## V. Argument for Principle 2: Pure Indexicals

As stated earlier, I believe this principle is uncontroversial. But I had best distinguish it from similar principles which are false. I am *not* claiming, as has been claimed for proper names, that indexicals lack anything that might be called 'descriptive meaning'. Indexicals, in general, have a rather easily statable descriptive meaning. But it is clear that this meaning is relevant only to determining a referent in a context of use and *not* to determining a relevant individual in a circumstance of evaluation. Let us return to the example in connection with the sentence (3) and the indexical 'I'. The bizarre result of taking the descriptive meaning of the indexical to be the propositional constituent is that what I said in uttering (3) would be true in a circumstance of evaluation if and only if the speaker (assuming there is one) of the circumstance does not exist in the circumstance. Nonsense! If *that* were the correct analysis, what I said could not be true. From which it follows that

It is impossible that I do not exist.

Here is another example to show that the descriptive meaning of an indexical may be entirely *inapplicable* in the circumstance of evaluation. When I say,

I wish I were not speaking now.

The circumstances desired do not involve contexts of *use* and *agents* who are not speaking. The *actual* context of use is used to determine the relevant individual: *me*—and time: *now*—and then we query the various circumstances of evaluation with respect to *that* individual and *that* time.

Here is another example, not of the inapplicability of the descriptive meaning to circumstances but of its irrelevance. Suppose I say at  $t_0$ , "It will soon be the case that all that is now beautiful is faded." Consider what was said in the subsentence,

All that is now beautiful is faded.

I wish to evaluate that content at some near future time  $t_1$ . What is the relevant time associated with the indexical 'now'? Is it the future time  $t_1$ ? No, it is  $t_0$ , of course: the time of the context of use.

See how rigidly the indexicals cling to the referent determined in the context of use:

- (4) It is possible that in Pakistan, in five years, only those who are actually here now are envied.

The point of (4) is that the circumstance, place, and time referred to by the indexicals 'actually', 'here', and 'now' are the circumstance, place, and time of the *context*, not a circumstance, place, and time determined by the modal, locational, and temporal operators within whose scope the indexicals lie.

It may be objected that this only shows that indexicals always take *primary* scope (in the sense of Russell's scope of a definite description). This objection attempts to relegate all direct reference to implicit use of the paradigm of the semantics of direct reference, the variable. Thus (4) is transformed into,

The actual circumstances, here, and now are such that it is possible that in Pakistan in five years only those who, in the first, are located at the second, during the third, are envied.

Although this may not be the most felicitous form of expression, its meaning and, in particular, its symbolization should be clear to those familiar with quantified intensional logics. The pronouns, 'the first', 'the second', and 'the third' are to be represented by distinct variables bound to existential quantifiers at the beginning and identified with 'the actual circumstance', 'here', and 'now' respectively.

- (5)  $(\exists w)(\exists p)(\exists t)[w = \text{the actual circumstance} \wedge p = \text{here} \wedge t = \text{now} \wedge \diamond \text{In Pakistan In five years } \forall x(x \text{ is envied} \rightarrow x \text{ is located at } p \text{ during } t \text{ in } w)]$

But such transformations, when thought of as representing the claim that indexicals take primary scope, do not provide an *alternative* to Principle 2, since we may still ask of an utterance of (5) in a context  $c$ , when evaluating it with respect to an arbitrary circumstance, to what do the indexicals 'actual', 'here', and 'now' refer. The answer, as always, is: the relevant features of the context  $c$ . [In fact, although (4) is equivalent to (5), neither indexicals nor quantification across intensional operators is dispensable in favor of the other.]

Perhaps enough has been said to establish the following.

(T1) *The descriptive meaning of a pure indexical determines the referent of the indexical with respect to a context of use but is either inapplicable or irrelevant to determining a referent with respect to a circumstance of evaluation.*

I hope that your intuition will agree with mine that it is for this reason that:

(T2) *When what was said in using a pure indexical in a context  $c$  is to be evaluated with respect to an arbitrary circumstance, the relevant object is always the referent of the indexical with respect to the context  $c$ .*

This is just a slightly elaborated version of Principle 2.

Before turning to true demonstratives, we will adopt some terminology.

## VI. Terminological Remarks

Principle 1 and Principle 2 taken together imply that sentences containing pure indexicals have two kinds of meaning.

### VI. (i) Content and Circumstance

What is said in using a given indexical in different contexts may be different. Thus if I say, today,

I was insulted yesterday

and you utter the same words tomorrow, what is said is different. If what we say differs in truth-value, that is enough to show that we say different things. But even if the truth-values were the same, it is clear that there are possible circumstances in which what I said would be true but what you said would be false. Thus we say different things.

Let us call this first kind of meaning—what is said—*content*. The content of a sentence in a given context is what has traditionally been called a proposition. Strawson, in noting that the sentence

The present king of France is bald

could be used on different occasions to make different statements, used 'statement' in a way similar to our use of *content of a sentence*. If we

wish to express the same content in different contexts, we may have to change indexicals. Frege, here using 'thought' for content of a sentence, expresses the point well.

If someone wants to say the same today as he expressed yesterday using the word 'today', he must replace this word with 'yesterday'. Although the thought is the same its verbal expression must be different so that the sense, which would otherwise be affected by the differing times of utterance, is readjusted.<sup>26</sup>

I take *content* as a notion applying not only to sentences taken in a context but to any meaningful part of speech taken in a context. Thus we can speak of the content of a definite description, an indexical, a predicate, etc. It is *contents* that are evaluated in circumstances of evaluation. If the content is a proposition (i.e., the content of a sentence taken in some context), the result of the evaluation will be a truth-value. The result of evaluating the content of a singular term at a circumstance will be an object (what I earlier called 'the relevant object'). In general, the result of evaluating the content of a well-formed expression  $\alpha$  at a circumstance will be an appropriate extension for  $\alpha$  (i.e., for a sentence, a truth-value; for a term, an individual; for an  $n$ -place predicate, a set of  $n$ -tuples of individuals, etc.). This suggests that we can represent a

<sup>26</sup>From "The Thought: A Logical Inquiry," *Mind* 65 (1956): 289–311. If Frege had only supplemented these comments with the observation that indexicals are devices of direct reference, the whole theory of indexicals would have been his. But his theory of meaning blinded him to this obvious point. Frege, I believe, mixed together the two kinds of meaning in what he called *Sinn*. A *thought* is, for him, the *Sinn* of a sentence, or perhaps we should say a *complete* sentence. *Sinn* is to contain both "the manner and context of presentation [of the denotation]," according to "Über Sinn und Bedeutung" (*Zeitschrift für Philosophie und philosophische Kritik* 100 (1892); trans. as "On Sense and Nominatum," in *Contemporary Readings in Logical Theory*, ed. Copi and Gould (Macmillan, 1967); mistrans. as "On Sense and Meaning," in Martinich, op. cit.). *Sinn* is first introduced to represent the cognitive significance of a sign, and thus to solve Frege's problem: how can ' $\alpha = \beta$ ' if true differ in cognitive significance from ' $\alpha = \alpha$ '. However, it also is taken to represent the truth-conditions or *content* (in our sense). Frege felt the pull of the two notions, which he reflects in some tortured passages about 'I' in "The Thought" (quoted below in XVII). If one says "Today is beautiful" on Tuesday and "Yesterday was beautiful" on Wednesday, one expresses the same thought according to the passage quoted. Yet one can clearly lose track of the days and not realize one is expressing the same thought. It seems then that thoughts are not appropriate bearers of cognitive significance. I return to this topic in XVII. A detailed examination of Frege on demonstratives is contained in John Perry's "Frege on Demonstratives," *Philosophical Review* 86 (1977): 474–97.

content by a function from circumstances of evaluation to an appropriate extension. Carnap called such functions *intensions*.

The representation is a handy one and I will often speak of contents in terms of it, but one should note that contents which are distinct but equivalent (i.e., share a value in all circumstances) are represented by the same intension. Among other things, this results in the loss of my distinction between terms which are devices of direct reference and descriptions which *turn out* to be rigid designators. (Recall the metaphysical paragraph of section IV.) I wanted the content of an indexical to be just the referent itself, but the intension of such a content will be a constant function. Use of representing intensions does not mean I am abandoning that idea—just ignoring it temporarily.

A *fixed content* is one represented by a constant function. All directly referential expressions (as well as all rigid designators) have a fixed content. [What I elsewhere call a *stable content*.]

Let us settle on *circumstances* for possible circumstances of evaluation. By this I mean both actual and counterfactual situations with respect to which it is appropriate to ask for the extensions of a given well-formed expression. A circumstance will usually include a possible state or history of the world, a time, and perhaps other features as well. The amount of information we require from a circumstance is linked to the degree of specificity of contents, and thus to the kinds of operators in the language.

Operators of the familiar kind treated in intensional logic (modal, temporal, etc.) operate on contents. (Since we represent contents by intensions, it is not surprising that intensional operators operate on contents.) Thus an appropriate extension for an intensional operator is a function from intensions to extensions.<sup>27</sup> A modal operator when applied to an intension will look at the behavior of the intension with respect to the possible state of the world feature of the circumstances of evaluation. A temporal operator will, similarly, be concerned with

<sup>27</sup>As we shall see, indexical operators such as "It is now the case that," "It is actually the case that," and "dthat" (the last takes a term rather than a sentence as argument) are also intensional operators. They differ from the familiar operators in only two ways: first, their extension (the function from intensions to extensions) depends on context, and second, they are directly referential (thus they have a fixed content). I shall argue below (in section VII: Monsters) that all operators that can be given an English reading are 'at most' intensional. Note that when discussing issues in terms of the formal representations of the model-theoretic semantics, I tend to speak in terms of intensions and intensional operators rather than contents and content operators.

the time of the circumstance. If we built the time of evaluation into the contents (thus removing time from the circumstances leaving only, say, a possible world history, and making contents *specific* as to time), it would make no sense to have temporal operators. To put the point another way, if *what is said* is thought of as incorporating reference to a specific time, or state of the world, or whatever, it is otiose to ask whether what is said would have been true at another time, in another state of the world, or whatever. Temporal operators applied to eternal sentences (those whose contents incorporate a specific time of evaluation) are redundant. Any intensional operators applied to *perfect* sentences (those whose contents incorporate specific values for all features of circumstances) are redundant.<sup>28</sup>

<sup>28</sup>The notion of redundancy involved could be made precise. When I speak of building the time of evaluation into contents, or making contents specific as to time, or taking what is said to incorporate reference to a specific time, what I have in mind is this. Given a sentence *S*: 'I am writing', in the present context *c*, which of the following should we take as the content: (i) the proposition that David Kaplan is writing at 10 A.M. on 3/26/77, or (ii) the 'proposition' that David Kaplan is writing? The proposition (i) is specific as to time, the 'proposition' (ii) [the scare quotes reflect my feeling that this is not the traditional notion of a proposition] is neutral with respect to time. If we take the content of *S* in *c* to be (ii), we can ask whether it would be true at times other than the time of *c*. Thus we think of the temporally neutral 'proposition' as changing its truth-value over time. Note that it is not just the noneternal sentence *S* that changes its truth-value over time, but the 'proposition' itself. Since the sentence *S* contains an indexical 'I', it will express different 'propositions' in different contexts. But since *S* contains no *temporal* indexical, the time of the context will not influence the 'proposition' expressed. An alternative [and more traditional] view is to say that the verb tense in *S* involves an implicit temporal indexical, so that *S* is understood as synonymous with *S'*: 'I am writing now'. If we take this point of view we will take the content of *S* in *c* to be (i). In this case *what is said* is eternal; it does not change its truth-value over time, although *S* will express different propositions at different times.

There are both technical and philosophical issues involved in choosing between (i) and (ii). Philosophically, we may ask why the temporal indexical should be taken to be implicit (making the proposition eternal) when no modal indexical is taken to be implicit. After all, we *could* understand *S* as synonymous with *S''*: 'I am actually writing now'. The content of *S''* in *c* is not only eternal, it is perfect. Its truth changes neither through time nor possibility. Is there some good philosophical reason for preferring contents which are neutral with respect to possibility but draw fixed values from the context for all other features of a possible circumstance whether or not the sentence contains an explicit indexical? (It may be that the traditional view was abetted by one of the delightful anomalies of the logic of indexicals, namely that *S*, *S'*, and *S''* are all logically equivalent! See Remark 3, p. 547.) Technically, we must note that intensional operators must, if they are not to be vacuous, operate on contents which are neutral with respect

What sorts of intensional operators to admit seems to me largely a matter of language engineering. It is a question of which features of what we intuitively think of as possible circumstances can be sufficiently well defined and isolated. If we wish to isolate location and regard it as a feature of possible circumstances we can introduce locational operators: 'Two miles north it is the case that', etc. Such operators can be iterated and can be mixed with modal and temporal operators. However, to make such operators interesting we must have contents which are locationally neutral. That is, it must be appropriate to ask if *what is said* would be true in Pakistan. (For example, 'It is raining' seems to be locationally as well as temporally and modally neutral.)

This functional notion of the content of a sentence in a context may not, because of the neutrality of content with respect to time and place, say, exactly correspond to the classical conception of a proposition. But the classical conception can be introduced by adding the demonstratives 'now' and 'here' to the sentence and taking the content of the result. I will continue to refer to the content of a sentence as a proposition, ignoring the classical use.

Before leaving the subject of circumstances of evaluation I should, perhaps, note that the mere attempt to show that an expression is directly referential requires that it be meaningful to ask of an individual in one circumstance whether and with what properties it exists in another circumstance. If such questions cannot be raised because they are regarded as metaphysically meaningless, the question of whether a particular expression is directly referential (or even, a rigid designator) cannot be raised. I have elsewhere referred to the view that such questions are meaningful as *haecceitism*, and I have described other metaphysical manifestations of this view.<sup>29</sup> I advocate this position, although I am

---

to the feature of circumstance the operator is interested in. Thus, for example, if we take the content of *S* to be (i), the application of a temporal operator to such a content would have no effect; the operator would be vacuous. Furthermore, if we do not wish the iteration of such operators to be vacuous, the content of the compound sentence containing the operator must again be neutral with respect to the relevant feature of circumstance. This is not to say that no such operator can have the effect of *fixing* the relevant feature and thus, in effect, rendering subsequent operations vacuous; indexical operators do just this. It is just that this must not be the general situation. A content must be the *kind* of entity that is subject to modification in the feature relevant to the operator. [The textual material to which this note is appended is too cryptic and should be rewritten.]

<sup>29</sup> "How to Russell a Frege-Church." The pronunciation is: "Hex-ee-i-tis-m." The epithet was suggested by Robert Adams. It is not an accident that it is derived from a demonstrative.

uncomfortable with some of its seeming consequences (for example, that the world might be in a state qualitatively exactly as it is, but with a permutation of individuals).

It is hard to see how one could think about the semantics of indexicals and modality without adopting such a view.

## VI. (ii) Character

The second kind of meaning, most prominent in the case of indexicals, is that which determines the content in varying contexts. The rule,

'I' refers to the speaker or writer

is a meaning rule of the second kind. The phrase 'the speaker or writer' is not supposed to be a complete description, nor it is supposed to refer to the speaker or writer of the word 'I'. (There are many such.) It refers to the speaker or writer of the relevant *occurrence* of the word 'I', that is, the agent of the context.

Unfortunately, as usually stated, these meaning rules are incomplete in that they do not explicitly specify that the indexical is directly referential, and thus do not completely determine the content in each context. I will return to this later.

Let us call the second kind of meaning, *character*. The character of an expression is set by linguistic conventions and, in turn, determines the content of the expression in every context.<sup>30</sup> Because character is what is set by linguistic conventions, it is natural to think of it as *meaning* in the sense of what is known by the competent language user.

Just as it was convenient to represent contents by functions from possible circumstances to extensions (Carnap's intentions), so it is convenient to represent characters by functions from possible contexts to contents. (As before we have the drawback that equivalent characters are identified.<sup>31</sup>) This gives us the following picture:

---

<sup>30</sup> This does not imply that if you know the character and are in first one and then another context, you can *decide* whether the contents are the same. I may twice use 'here' on separate occasions and not recognize that the place is the same, or twice hear 'I' and not know if the content is the same. What I do know is this: if it was the same person speaking, then the content was the same. [More on this epistemological stuff later.]

<sup>31</sup> I am, at this stage, deliberately ignoring Kripke's theory of proper names in order to see whether the revisions in Fregean semantical theory, which seem plainly required to accommodate indexicals (this is the 'obviousness' of my theory), can throw any light on it. Here we assume that aside from indexicals, Frege's theory

Character: Contexts  $\Rightarrow$  Contents

Content: Circumstances  $\Rightarrow$  Extensions

or, in more familiar language,

Meaning + Context  $\Rightarrow$  Intension

Intension + Possible World  $\Rightarrow$  Extension

Indexicals have a *context-sensitive* character. It is characteristic of an indexical that its content varies with context. Nonindexicals have a *fixed* character. The same content is invoked in all contexts. This content will typically be sensitive to circumstances, that is, the non-indexicals are typically not rigid designators but will vary in extension from circumstance to circumstance. Eternal sentences are generally good examples of expressions with a fixed character.

All persons alive in 1977 will have died by 2077

expresses the same proposition no matter when said, by whom, or under what circumstances. The truth-value of that proposition may, of course, vary with possible circumstances, but the character is fixed. Sentences with fixed character are very useful to those wishing to leave historical records.

Now that we have two kinds of meaning in addition to extension, Frege's principle of intensional interchange<sup>32</sup> becomes two principles:

is correct, roughly, that words and phrases have a kind of descriptive meaning or sense which at one and the same time constitutes their cognitive significance and their conditions of applicability.

Kripke says repeatedly in *Naming and Necessity* that he is only providing a picture of how proper names refer and that he does not have an exact theory. His picture yields some startling results. In the case of indexicals we do have a rather precise theory, which avoids the difficulty of specifying a chain of communication and which yields many analogous results. In facing the vastly more difficult problems associated with a theory of reference for proper names, the theory of indexicals may prove useful; if only to show—as I believe—that proper names are not indexicals and have no meaning in the sense in which indexicals have meaning (namely a 'cognitive content' which fixes the references in all contexts). [The issues that arise, involving token reflexives, homonymous words with distinct character, and homonymous token reflexives with the same character are best saved for later—much later.]

<sup>32</sup>See §28 of Rudolf Carnap's *Meaning and Necessity* (Chicago: University of Chicago Press, 1947).

(F1) The character of the whole is a function of the character of the parts. That is, if two compound well-formed expressions differ only with respect to components which have the same Character, then the Character of the compounds is the same.

(F2) The Content of the whole is a function of the Content of the parts. That is, if two compound well-formed expressions, each set in (possibly different) contexts differ only with respect to components which *when taken in their respective contexts* have the same content, then the content of the two compounds *each taken in its own context* is the same.

It is the second principle that accounts for the often noted fact that speakers in different contexts can say the same thing by switching indexicals. (And indeed they often *must* switch indexicals to do so.) Frege illustrated this point with respect to 'today' and 'yesterday' in "The Thought." (But note that his treatment of 'I' suggests that he does not believe that utterances of 'I' and 'you' could be similarly related!)

Earlier, in my metaphysical phase, I suggested that we should think of the content of an indexical as being just the referent itself, and I resented the fact that the representation of contents as intensions forced us to regard such contents as constant functions. A similar remark applies here. If we are not overly concerned with standardized representations (which certainly have their value for model-theoretic investigations) we might be inclined to say that the character of an indexical-free word or phrase just *is* its (constant) content.

## VII. Earlier Attempts: Index Theory

The following picture seems to emerge. The meaning (character) of an indexical is a function from contexts to extensions (substituting for fixed contents). The meaning (content, substituting for fixed characters) of a nonindexical is a function from circumstances to extensions. From this point of view it may appear that the addition of indexicals requires no new *logic*, no sharp distinction between contexts and circumstances, just the addition of some special new *features* ('contextual' features) to the circumstances of evaluation. (For example, an *agent* to provide an interpretation for 'I'.) Thus an enlarged view of intension is derived. The intension of an expression is a function from certain factors to the extension of the expression (with respect to those factors). Originally such factors were simply possible states of the world, but as it was noticed

that the so-called tense operators exhibited a structure highly analogous to that of the modal operators the factors with respect to which an extension was to be determined were enlarged to include moments of time. When it was noticed that contextual factors were required to determine the extension of sentences containing indexicals, a still more general notion was developed and called an "index." The extension of an expression was to be determined with respect to an index. The intension of an expression was that function which assigned to every index, the extension at that index.

The above example supplies us with a statement whose truth-value is not constant but varies as a function of  $i \in I$ . This situation is easily appreciated in the context of time-dependent statements; that is, in the case where  $I$  represents the instant of time. Obviously the same statement can be true at one moment and false at another. For more general situations one must not think of the  $i \in I$  as anything as simple as instants of time or even possible worlds. In general we will have

$$i = (w, t, p, a, \dots)$$

where the index  $i$  has many *coordinates*: for example,  $w$  is a *world*,  $t$  is a *time*,  $p = (x, y, z)$  is a (3-dimensional) *position* in the world,  $a$  is an *agent*, etc. All these coordinates can be varied, possibly independently, and thus affect the truth-values of statements which have indirect references to these coordinates. [From the *Advice* of a prominent logician.]

A sentence  $\phi$  was taken to be logically true if true at every index (in every 'structure'), and  $\Box\phi$  was taken to be true at a given index (in a given structure) just in case  $\phi$  was true at every index (in that structure). Thus the familiar principle of modal generalization: if  $\models \phi$ , then  $\models \Box\phi$ , is validated.

This view, in its treatment of indexicals, was technically wrong and, more importantly, conceptually misguided.

Consider the sentence

(6) I am here now.

It is obvious that for many choices of index—i.e., for many quadruples  $\langle w, x, p, t \rangle$  where  $w$  is a possible world history,  $x$  is a person,  $p$  is a place,

and  $t$  is a time—(6) will be false. In fact, (6) is true only with respect to those indices  $\langle w, x, p, t \rangle$  which are such that in the world history  $w$ ,  $x$  is located at  $p$  at the time  $t$ . Thus (6) fares about on a par with

(7) David Kaplan is in Portland on 26 March 1977.

(7) is empirical, and so is (6).

But here we have missed something essential to our understanding of indexicals. Intuitively, (6) is deeply, and in some sense, which we will shortly make precise, universally, true. One need only understand the meaning of (6) to know that it cannot be uttered falsely. No such guarantees apply to (7). A *Logic of Indexicals* which does not reflect this intuitive difference between (6) and (7) has bypassed something essential to the logic of indexicals.

What has gone wrong? We have ignored the special relationship between 'I', 'here', and 'now'. Here is a proposed correction. Let the class of indices be narrowed to include only the *proper* ones—namely, those  $\langle w, x, p, t \rangle$  such that in the world  $w$ ,  $x$  is located at  $p$  at the time  $t$ . Such a move may have been intended originally since improper indices are like impossible worlds; no such contexts *could* exist and thus there is no interest in evaluating the extensions of expressions with respect to them. Our reform has the consequence that (6) comes out, correctly, to be logically true. Now consider

(8)  $\Box$  I am here now.

Since the contained sentence (namely (6)) is true at every proper index, (8) also is true at every proper index and thus also is logically true. (As would be expected by the aforementioned principle of modal generalization.)

But (8) should not be *logically* true, since it is false. It is certainly *not* necessary that I be here now. But for several contingencies, I would be working in my garden now, or even delivering this paper in a location outside of Portland.

The difficulty, here, is the attempt to assimilate the role of a *context* to that of a *circumstance*. The indices  $\langle w, x, p, t \rangle$  that represent contexts must be proper in order that (6) be a truth of the logic of indexicals, but the indices that represent circumstances must include improper ones in order that (8) *not* be a logical truth.

If one wishes to stay with this sort of index theory and blur the conceptual difference between context and circumstance, the minimal requirement is a system of *double* indexing, one index for context and

another for circumstance. It is surprising, looking back, that we (for I was among the early index theorists) did not immediately see that double indexing was required, for in 1967, at UCLA, Hans Kamp had reported his work on 'now'<sup>33</sup> in which he had shown that double indexing was required to properly accommodate temporal indexicals along with the usual temporal operators. But it was *four* years before it was realized that this was a general requirement for (and, in a sense, the key to) a logic of indexicals.

However, mere double indexing, without a clear conceptual understanding of what each index stands for, is still not enough to avoid all pitfalls.

### VIII. Monsters Begat by Elegance

My liberality with respect to operators on content, i.e., intensional operators (any feature of the circumstances of evaluation that can be well defined and isolated) does not extend to operators which attempt to operate on character. Are there such operators as 'In some contexts it is true that', which when prefixed to a sentence yields a truth if and only if in some context the contained *sentence* (not the content expressed by it) expresses a content that is true in the circumstances of that context? Let us try it:

(9) In some contexts it is true that I am not tired now.

For (9) to be true in the present context it suffices that some agent of some context not be tired at the time of that context. (9), so interpreted, has nothing to do with me or the present moment. But this violates Principle 2! Principle 2 can also be expressed in more theory laden way by saying that indexicals always take primary scope. If this is true—and it is—then no operator can control the character of the indexicals within its scope, because they will simply leap out of its scope to the front of the operator. I am not saying we could not construct a language with such operators, just that English is not one.<sup>34</sup> And such operators *could not be added to it*.

There *is* a way to control an indexical, to keep it from taking primary scope, and even to refer it to another context (this amounts to changing its character). Use quotation marks. If we *mention* the indexical rather

than *use* it, we can, of course, operate directly on it. Carnap once pointed out to me how important the difference between direct and indirect quotation is in

Otto said "I am a fool."

Otto said that I am a fool.

Operators like 'In some contexts it is true that', which attempt to meddle with character, I call *monsters*. I claim that none can be expressed in English (without sneaking in a quotation device). If they stay in the metalanguage and confine their attention to sentences as in

In some contexts "I am not tired now" is true

they are rendered harmless and can even do socially useful work (as does, 'is valid' [see below]).

I have gone on at perhaps excessive length about monsters because they have recently been begat by elegance. In a specific application of the theory of indexicals there will be just certain salient features of a circumstance of evaluation. So we may represent circumstances by indexed sets of features. This is typical of the model-theoretic way. As already indicated, all the features of a circumstance will generally be required as aspects of a context, and the aspects of a context may all be features of a circumstance. If not, a little ingenuity may make it so.<sup>35</sup>

<sup>35</sup>Recall that in a particular formal theory the features of a circumstance must include all elements with respect to which there are content operators, and the aspects of a context must include all elements with respect to which there are indexicals. Thus, a language with both the usual modal operators ' $\Diamond$ ', ' $\Box$ ', and an indexical modal operator 'It is actually the case that' will contain a possible world history feature in its circumstances as well as an analogous aspect in its contexts. If a circumstance is an aspect of a context, as seems necessary for the definition of truth, then we only need worry about aspects of contexts that are not features of circumstances. The most prominent of these is the *agent* of the context, required to interpret the indexical 'I'. In order to supply a corresponding nonvacuous feature to circumstances we must treat contents in such a way that we can ask whether they are true for various agents. (Not *characters* mind you, but contents.) This can be done by representing the agent by a *neutral*—a term which plays the syntactical role of 'I' but gets an interpretation only with respect to a circumstance. Let  $a$  be a special variable that is not subject to quantification and let  $b$  be a variable not in the language. Our variable  $a$  is the neutral. We wish to introduce content operators which affect the agent place and which can be iterated. Let  $R$  be a relation between individuals, for example ' $aRb$ ' for ' $b$  is an uncle of  $a$ '. Then we may interpret the operator  $O^R\phi$  as  $(\exists b)[aRb \wedge (\exists a)(b = a \wedge \phi)]$ . If  $\phi$  is ' $a$  walks',  $O^R\phi$  comes to ' $a$  uncle of  $a$  walks'. The indexical 'I' can be represented by an operator  $O^I$  for which ' $aRb$ ' is just ' $I=b$ '. The result should be that  $O^I\phi$  is equivalent to replacing the neutral  $a$  by the indexical 'I'.

<sup>33</sup>Published in 1971 as "Formal Properties of 'Now'," *Theoria*.

<sup>34</sup>Thomason alleges a counterinstance: 'Never put off until tomorrow what you can do today'. What should one say about this?



We could then represent contexts by the same indexed sets we use to represent circumstances, and instead of having a *logic of contexts and circumstances* we have simply a *two-dimensional logic of indexed sets*. This is algebraically very neat and it permits a very simple and elegant description of certain important classes of characters (for example, those which are true at every pair  $\langle i, i \rangle$ , though the special significance of the set is somehow diminished in the abstract formulation).<sup>36</sup> But it also permits a simple and elegant introduction of many operators which are monsters. In abstracting from the distinct conceptual roles played by contexts of use and circumstances of evaluation the special logic of indexicals has been obscured. Of course restrictions can be put on the two-dimensional logic to exorcise the monsters, but to do so would be to give up the mathematical advantages of that formulation.<sup>37</sup>

### IX. Argument for Principle 2: True Demonstratives

I return now to the argument that all indexicals are directly referential. Suppose I point at Paul and say,

He now lives in Princeton, New Jersey.

Call *what I said*—i.e., the content of my utterance, the proposition expressed—'Pat'. Is Pat true or false? True! Suppose that unbeknownst to me, Paul had moved to Santa Monica last week. Would Pat have then been true or false? False! Now, the tricky case: Suppose that Paul and Charles had each disguised themselves as the other and had switched places. If that had happened, *and* I had uttered as I did, then the proposition I *would have* expressed would have been false. But in that possible context the proposition I *would have* expressed is not Pat. That is easy to see because the proposition I *would have* expressed, had I pointed to Charles instead of Paul—call this proposition 'Mike'—not only *would have* been false but actually is false. Pat, I would claim, would still be true in the circumstances of the envisaged possible con-

text provided that Paul—in whatever costume he appeared—were still residing in Princeton.

### IX. (i) The Arguments

I am arguing that in order to determine what the truth-value of a proposition expressed by a sentence containing a demonstrative *would be* under other possible circumstances, the relevant individual is not the individual that *would have* been demonstrated had those circumstances obtained and the demonstration been set in a context of those circumstances, but rather the individual demonstrated in the context which *did* generate the proposition being evaluated. As I have already noted, it is characteristic of sentences containing demonstratives—or, for that matter, any indexical—that they may express different propositions in different contexts. We must be wary of confusing the proposition that would have been expressed by a similar utterance in a slightly different context—say, one in which the demonstratum is changed—with the proposition that was actually expressed. If we keep this distinction in mind—i.e., we distinguish Pat and Mike—we are less likely to confuse what the truth-value of the proposition *actually* expressed would have been under some possible circumstances with what the truth-value of the proposition that *would have been* expressed would have been under those circumstances.

When we consider the vast array of possible circumstances with respect to which we might inquire into the truth of a proposition expressed in some context *c* by an utterance *u*, it quickly becomes apparent that only a small fraction of these circumstances will involve an utterance of the same sentence in a similar context, and that there must be a way of evaluating the truth-value of propositions expressed using demonstratives in counterfactual circumstances in which no demonstrations are taking place and no individual has the exact characteristics exploited in the demonstration. Surely, it is irrelevant to determining whether what I said would be true or not in some counterfactual circumstance, whether Paul, or anyone for that matter, *looked* as he does now. All that would be relevant is *where he lives*. Therefore,

(T3) the relevant features of the demonstratum *qua demonstratum* (compare, the relevant features of the  $x$   $Fx$  *qua the x Fx*)—namely, that the speaker is pointing at it, that it has a certain appearance, is presented in a certain way—cannot be the essential characteristics used to identify the relevant individual in counterfactual situations.

<sup>36</sup>See, for example, Krister Segerberg, "Two-dimensional Modal Logic," *Journal of Philosophical Logic* 2 (1973): 77–96. Segerberg does metamathematical work in his article and makes no special philosophical claims about its significance. That has been done by others.

<sup>37</sup>There is one other difficulty in identifying the class of contexts with the class of circumstances. The special relationship between the indexicals 'I', 'here', 'now' seems to require that the agent of a context be at the location of the context during the time of the context. But this restriction is not plausible for arbitrary circumstances. It appears that this approach will have difficulty in avoiding the problems of (6) and (8) (section VII).

These two arguments: the distinction between Pat and Mike, and consideration of counterfactual situations in which no demonstration occurs, are offered to support the view that demonstratives are devices of direct reference (rigid designators, if you will) and, by contrast, to reject a Fregean theory of demonstratives.

### IX. (ii) The Fregean Theory of Demonstrations

In order to develop the latter theory, in contrast to my own, we turn first to a portion of the Fregean theory which I accept: the Fregean theory of demonstrations.

As you know, for a Fregean the paradigm of a meaningful expression is the definite description, which picks out or denotes an individual, a unique individual, satisfying a condition *s*. The individual is called the *denotation* of the definite description and the condition *s* we may identify with the *sense* of the definite description. Since a given individual may uniquely satisfy several distinct conditions, definite descriptions with distinct senses may have the same denotation. And since some conditions may be uniquely satisfied by no individual, a definite description may have a sense but no denotation. The condition by means of which a definite description picks out its denotation is the *manner of presentation* of the denotation by the definite description.

The Fregean theory of demonstratives claims, correctly I believe, that the analogy between descriptions (short for 'definite descriptions') and demonstrations is close enough to provide a sense and denotation analysis of the 'meaning' of a demonstration. The denotation is the demonstratum (that which is demonstrated), and it seems quite natural to regard each demonstration as presenting its demonstratum in a particular manner, which we may regard as the sense of the demonstration. The same individual could be demonstrated by demonstrations so different in manner of presentation that it would be informative to a competent auditor-observer to be told that the demonstrata were one. For example, it might be informative to you for me to tell you that

That [pointing to Venus in the morning sky] is identical with that [pointing to Venus in the evening sky].

(I would, of course, have to speak very slowly.) The two demonstrations—call the first one 'Phos' and the second one 'Hes'—which accompanied the two occurrences of the demonstrative expression 'that' have

the same demonstratum but distinct manners of presentation. It is this difference between the sense of Hes and the sense of Phos that accounts, the Fregean claims, for the informativeness of the assertion.

It is possible, to pursue the analogy, for a demonstration to have no demonstratum. This can arise in several ways: through hallucination, through carelessness (not noticing, in the darkened room, that the subject had jumped off the demonstration platform a few moments before the lecture began), through a sortal conflict (using the demonstrative phrase 'that *F*', where *F* is a common noun phrase, while demonstrating something which is not an *F*), and in other ways.

Even Donnellans's important distinction between referential and attributive uses of definite descriptions seems to fit, equally comfortably, the case of demonstrations.<sup>38</sup>

The Fregean hypostatizes demonstrations in such a way that it is appropriate to ask of a given demonstration, say Phos, what *would* it have demonstrated under various counterfactual circumstances. Phos and Hes might have demonstrated distinct individuals.<sup>39</sup>

We should not allow our enthusiasm for analogy to overwhelm judgment in this case. There are some relevant respects in which descriptions and demonstrations are disanalogous. First, as David Lewis has pointed out, demonstrations do not have a syntax, a fixed formal structure in terms of whose elements we might try to define, either directly or recursively, the notion of sense.<sup>40</sup> Second, to different audiences (for example, the speaker, those sitting in front of the demonstration platform, and those sitting behind the demonstration platform) the same demonstration may have different senses. Or perhaps we should say that a single performance may involve distinct demonstrations from the perspective of distinct audiences. ("Exactly like proper names!" says the Fregean, "as long as the demonstratum remains the same, these fluctuations in sense are tolerable. But they should be avoided in the system

<sup>38</sup>I have written elsewhere, in appendices VII and VIII of "Bob and Carol and Ted and Alice," of these matters and won't pursue the topic now.

<sup>39</sup>It could then be proposed that demonstrations be individuated by the principle:  $d_1 = d_2$  if and only if, for all appropriate circumstances *c*, the demonstratum of  $d_1$  in *c* = the demonstratum of  $d_2$  in *c*. An alternative principle of individuation is that the same demonstration is being performed in two different contexts if the standard audience can't determine, from the demonstration alone, whether the contexts are distinct or identical. This makes the individuation of demonstrations more epistemological than the metaphysical proposal above.

<sup>40</sup>Although recent work on computer perception has attempted to identify a syntax of pictures. See P. Suppes and Rottmayer, "Automata," in *Handbook of Perception*, vol. 1 (New York: Academic Press, 1974).

of a demonstrative science and should not appear in a perfect vehicle of communication.”)

### IX. (iii) The Fregean Theory of Demonstratives

Let us accept, tentatively and cautiously, the Fregean theory of demonstrations, and turn now to the Fregean theory of demonstratives.<sup>41</sup>

According to the Fregean theory of demonstratives, an occurrence of a demonstrative expression functions rather like a place-holder for the associated demonstration. The sense of a sentence containing demonstratives is to be the result of replacing each demonstrative by a constant whose sense is given as the sense of the associated demonstration. An important aim of the Fregean theory is, of course, to solve Frege's problem. And it does that quite neatly. You recall that the Fregean accounted for the informativeness of

That [Hes] = that [Phos]

in terms of the distinct senses of Hes and Phos. Now we see that the senses of the two occurrences of 'that' are identified with these two distinct senses so that the ultimate solution is exactly like that given by Frege originally. The sense of the left 'that' differs from the sense of the right 'that'.

### IX. (iv) Argument Against the Fregean Theory of Demonstratives

Let us return now to our original example:

He [Delta] now lives in Princeton, New Jersey

where 'Delta' is the name of the relevant demonstration. I assume that in the possible circumstances described earlier, Paul and Charles having disguised themselves as each other, Delta would have demonstrated Charles. Therefore, according to the Fregean theory, the proposition I just expressed, Pat, would have been false under the counterfactual circumstances of the switch. But this, as argued earlier, is wrong. Therefore, the Fregean theory of demonstratives though it nicely solves Frege's problem, is simply incorrect in associating propositions with utterances.

Let me recapitulate. We compared two theories as to the proposition expressed by a sentence containing a demonstrative along with an asso-

ciated demonstration. Both theories allow that the demonstration can be regarded as having both a sense and a demonstratum. My theory, the direct reference theory, claims that in assessing the proposition in counterfactual circumstances it is the actual demonstratum—in the example, Paul—that is the relevant individual. The Fregean theory claims that the proposition is to be construed as if the sense of the demonstration were the sense of the demonstrative. Thus, in counterfactual situations it is the individual that *would* have been demonstrated that is the relevant individual. According to the direct reference theory, demonstratives are rigid designators. According to the Fregean theory, their denotation varies in different counterfactual circumstances as the demonstrata of the associated demonstration would vary in those circumstances.

The earlier distinction between Pat and Mike, and the discussion of counterfactual circumstances in which, as we would now put it, the demonstration would have demonstrated nothing, argue that with respect to the problem of associating propositions with utterances the direct reference theory is correct and the Fregean theory is wrong.

I have carefully avoided arguing for the direct reference theory by using modal or subjunctive sentences for fear the Fregean would claim that the peculiarity of demonstratives is not that they are rigid designators but that they always take primary scope. If I had argued only on the basis of our intuitions as to the truth-value of

If Charles and Paul had changed chairs, then he (Delta)  
would not now be living in Princeton

such a scope interpretation could be claimed. But I didn't.

The perceptive Fregeans among you will have noted that I have said nothing about how Frege's problem fares under a direct reference theory of demonstratives. And indeed, if 'that' accompanied by a demonstration is a rigid designator for the demonstratum, then

that (Hes) = that (Phos)

looks like two rigid designators designating the same thing. Uh Oh! I will return to this in my Epistemological Remarks (section XVII).

<sup>41</sup>The Fregean theory of demonstrations is not a part of my obvious and uncontroversial theory of indexicals. On the contrary, it has the fascination of the speculative.

## X. Fixing the Reference vs. Supplying a Synonym<sup>42</sup>

The Fregean is to be forgiven. He has made a most natural mistake. Perhaps he thought as follows: If I point at someone and say 'he', that occurrence of 'he' must refer to the male at whom I am now pointing. It does! So far, so good. Therefore, the Fregean reasons, since 'he' (in its demonstrative sense) means the same as 'the male at whom I am now pointing' and since the denotation of the latter varies with circumstances the denotation of the former must also. But this is wrong. Simply because it is a rule of the language that 'he' *refers* to the male at whom I am now pointing (or, whom I am now demonstrating, to be more general), it does not follow that any synonymy is thereby established. In fact, this is one of those cases in which—to use Kripke's excellent idiom—the rule simply tells us how to *fix the reference* but does not supply a synonym.

Consider the proposition I express with the utterance

He [Delta] is the male at whom I am now pointing.

Call that proposition 'Sean'. Now Sean is certainly true. We know from the rules of the language that any utterance of that form must express a true proposition. In fact we would be justified in calling the *sentence*

He is the male at whom I am now pointing.

almost analytic. ('Almost' because of the hypothesis that the demonstrative is *proper*—that I am pointing at a unique male—is needed.)

But is Sean necessary? Certainly not, I might have pointed at someone else.

This kind of mistake—to confuse a semantical rule which tells how to fix the reference to a directly referential term with a rule which supplies a synonym—is easy to make. Since semantics must supply a meaning, in the sense of content (as I call it), for expressions, one thinks naturally that whatever way the referent of an expression is given by the semantical rules, that *way* must stand for the content of the expression. (Church [or was it Carnap?] says as much, explicitly.) This hypothesis

<sup>42</sup>I use Kripke's terminology to expound the important distinction he introduces in *Naming and Necessity* for descriptive meaning that may be associated with a proper name. As in several other cases of such parallels between proper names and indexicals, the distinction, and its associated argument, seems more obvious when applied to indexicals.

seems especially plausible, when, as is typical of indexicals,

the semantical rule which fixes the reference seems to exhaust our knowledge of the meaning of the expression.

## X. (i) Reichenbach on Token Reflexives

It was from such a perspective, I believe, that Reichenbach built his ingenious theory of indexicals. Reichenbach called such expressions 'token-reflexive words' in accordance with his theory. He writes as follows:

We saw that most individual-descriptions are constructed by reference to other individuals. Among these there is a class of descriptions in which the individual referred to is the act of speaking. We have special words to indicate this reference; such words are 'I', 'you', 'here', 'now', 'this'. Of the same sort are the tenses of verbs, since they determine time by reference to the time when the words are uttered. To understand the function of these words we have to make use of the distinction between *token* and *symbol*, 'token' meaning the individual sign, and 'symbol' meaning the class of similar tokens (cf. §2). Words and sentences are symbols. The words under consideration are words which refer to the corresponding token used in an individual act of speech, or writing; they may therefore be called *token-reflexive* words.

It is easily seen that all these words can be defined in terms of the phrase 'this token'. The word 'I', for instance, means the same as 'the person who utters this token'; 'now' means the same as 'the time at which this token was uttered'; 'this table' means the same as 'the table pointed to by a gesture accompanying this token'. We therefore need inquire only into the meaning of the phrase 'this token'.<sup>43</sup>

But is it true, for example, that

- (10) 'I' means the same as 'the person who utters this token' ?

It is certainly true that

I am the person who utters this token.

<sup>43</sup>H. Reichenbach, *Elements of Symbolic Logic* (New York: Macmillan, 1947), p. 284.

But if (10) correctly asserted a synonymy, then it would be true that

(11) If no one were to utter this token, I would not exist.

Beliefs such as (11) could make one a compulsive talker.

## XI. The Meaning of Indexicals

In order to correctly and more explicitly state the semantical rule which the dictionary attempts to capture by the entry

I: the person who is speaking or writing

we would have to develop our semantical theory—the semantics of direct reference—and then state that

(D1) 'I' is an indexical, different utterances of which may have different contents

(D3) 'I' is, in each of its utterances, directly referential

(D2) In each of its utterances, 'I' refers to the person who utters it.

We have seen errors in the Fregean analysis of demonstratives and in Reichenbach's analysis of indexicals, all of which stemmed from failure to realize that these words are directly referential. When we say that a word is directly referential are we saying that its meaning *is* its reference (its only meaning is its reference, its meaning is nothing more than its reference)? Certainly not.<sup>44</sup> Insofar as meaning is given by the rules of a language and is what is known by competent speakers, I would be more inclined to say in the case of directly referential words and phrases that their reference is *no* part of their meaning. The meaning of the word 'I' does not change when different persons use it. The meaning of 'I' is given by the rules (D1), (D2), and (D3) above.

<sup>44</sup>We see here a drawback to the terminology 'direct reference'. It suggests falsely that the reference is not mediated by a meaning, which it is. The meaning (character) is directly associated, by convention, with the word. The meaning determines the referent; and the referent determines the content. It is this to which I alluded in the parenthetical remark following the picture on page 486. Note, however, that the kind of descriptive meaning involved in giving the character of indexicals like 'I', 'now', etc., is, because of the focus on context rather than circumstance, unlike that traditionally thought of as Fregean sense. It is the idea that the referent determines the content—that, contra Frege, there is a road back—that I wish to capture. This is the importance of Principle 2.

Meanings tell us how the content of a word or phrase is determined by the context of use. Thus the meaning of a word or phrase is what I have called its *character*. (Words and phrases with no indexical element express the same content in every context; they have a fixed character.) To supply a synonym for a word or phrase is to find another with the same *character*; finding another with the same *content* in a particular context certainly won't do. The content of 'I' used by me may be identical with the content of 'you' used by you. This doesn't make 'I' and 'you' synonyms. Frege noticed that if one wishes to say again what one said yesterday using 'today', today one must use 'yesterday'. (Incidentally the relevant passage, quoted on page 501, propounds what I take to be a direct reference theory of the indexicals 'today' and 'yesterday'.) But 'today' and 'yesterday' are not synonyms. For two words or phrases to be synonyms, they must have the same content in every context. In general, for indexicals, it is not possible to find synonyms. This is because indexicals are directly referential, and the compound phrases which can be used to give their reference ('the person who is speaking', 'the individual being demonstrated', etc.) are not.

## XII. Dthat<sup>45</sup>

It would be useful to have a way of converting an arbitrary singular term into one which is directly referential.

Recall that we earlier regarded demonstrations, which are required to 'complete' demonstratives, as a kind of description. The demonstrative was then treated as a directly referential term whose referent was the demonstratum of the associated demonstration.

Now why not regard descriptions as a kind of demonstration, and introduce a special demonstrative which requires completion by a description and which is treated as a directly referential term whose referent is the denotation of the associated description? Why not? Why not indeed! I have done so, and I write it thus:

dthat[ $\alpha$ ]

where  $\alpha$  is any description, or, more generally, any singular term. 'Dthat' is simply the demonstrative 'that' with the following singular term func-

<sup>45</sup>Pronunciation note on 'dthat'. The word is not pronounced dee-that or duh-that. It has only one syllable. Although articulated differently from 'that' (the tongue begins behind the teeth), the sounds are virtually indistinguishable to all but native speakers.

tioning as its demonstration. (Unless you hold a Fregean theory of demonstratives, in which case its meaning is as stipulated above.)

Now we can come much closer to providing genuine synonyms.

'I' means the same as 'dthat [the person who utters this token]'.

(The fact that this alleged synonymy is cast in the theory of utterances rather than occurrences introduces some subtle complications, which have been discussed by Reichenbach.)

### XIII. Contexts, Truth, and Logical Truth

I wish, in this section, to contrast an *occurrence* of a well-formed expression (my *technical* term for the combination of an expression and a context) with an *utterance* of an expression.

There are several arguments for my notion, but the main one is from Remark 1 on the Logic of Demonstratives (section XIX below): I have sometimes said that the content of a sentence in a context is, roughly, the proposition the sentence would express if uttered in that context. This description is not quite accurate on two counts. First, it is important to distinguish an *utterance* from a *sentence-in-a-context*. The former notion is from the theory of speech acts, the latter from semantics. Utterances take time, and utterances of distinct sentences cannot be simultaneous (i.e., in the same context). But in order to develop a logic of demonstratives we must be able to evaluate several premises and a conclusion all in the same context. We do not want arguments involving indexicals to become valid simply because there is no possible context in which all the premises are uttered, and thus no possible context in which all are uttered truthfully.

Since the content of an occurrence of a sentence containing indexicals depends on the context, the notion of *truth* must be relativized to a context.

If  $c$  is a context, then an occurrence of  $\phi$  in  $c$  is true iff the content expressed by  $\phi$  in this context is true when evaluated with respect to the circumstance of the context.

We see from the notion of truth that among other aspects of a context must be a possible circumstance. Every context occurs in a particular circumstance, and there are demonstratives such as 'actual' which refer to that circumstance.

If you try out the notion of truth on a few examples, you will see that it is correct. If I now utter a sentence, I will have uttered a truth just in case *what I said*, the content, is true in *these* circumstances.

As is now common for intensional logics, we provide for the notion of a *structure*, comprising a family of circumstances. Each such structure will determine a set of possible contexts. Truth in a structure, is truth in every possible context of the structure. Logical truth is truth in every structure.

### XIV. Summary of Findings (so far): Pure Indexicals

Let me try now to summarize my findings regarding the semantics of demonstratives and other indexicals. First, let us consider the non-demonstrative indexicals such as 'I', 'here' (in its nondemonstrative sense), 'now', 'today', 'yesterday', etc. In the case of these words, the linguistic conventions which constitute *meaning* consist of rules specifying the referent of a given *occurrence* of the word (we might say, a given token, or even utterance, of the word, if we are willing to be somewhat less abstract) in terms of various features of the context of the occurrence. Although these rules fix the referent and, in a very special sense, might be said to define the indexical, the way in which the rules are given does not provide a synonym for the indexical. The rules tell us for any possible occurrence of the indexical what the referent would be, but they do *not* constitute the content of such an occurrence. Indexicals are directly referential. The rules tell us what it is that is referred to. Thus, they *determine* the content (the propositional constituent) for a particular occurrence of an indexical. But they are not a *part* of the content (they constitute no part of the propositional constituent). In order to keep clear on a topic where ambiguities constantly threaten, I have introduced two technical terms: *content* and *character* for the two kinds of meaning (in addition to extension) I associate with indexicals. Distinct occurrences of an indexical (in distinct contexts) may not only have distinct referents, they may have distinct meanings in the sense of *content*. If I say "I am tired today" today and Montgomery Furth says "I am tired today" tomorrow, our utterances have different contents in that the factors which are relevant to determining the truth-value of what Furth said in both actual and counterfactual circumstances are quite different from the factors which are relevant to determining the truth-value of what I said. Our two utterances are as different in content as are the sentences "David Kaplan is tired on 26 March 1977" and

"Montgomery Furth is tired on 27 March 1977." But there is another sense of meaning in which, absent lexical or syntactical ambiguities, two occurrences of the *same* word or phrase *must* mean the same. (Otherwise how could we learn and communicate with language?) This sense of meaning—which I call *character*—is what determines the content of an occurrence of a word or phrase in a given context. For indexicals, the rules of language constitute the meaning in the sense of *character*. As normally expressed, in dictionaries and the like, these rules are incomplete in that, by omitting to mention that indexicals are directly referential, they fail to specify the full content of an occurrence of an indexical.

Three important features to keep in mind about these two kinds of meaning are:

1. Character applies only to words and phrases as types, content to occurrences of words and phrases in contexts.
2. Occurrences of two phrases can agree in content although the phrases differ in character, and two phrases can agree in character but differ in content in distinct contexts.
3. The relationship of character to content is something like that traditionally regarded as the relationship of sense to denotation, character is a way of presenting content.

## XV. Further Details: Demonstratives and Demonstrations

Let me turn now to the demonstratives proper, those expressions which must be associated with a demonstration in order to determine a referent. In addition to the pure demonstratives 'that' and 'this' there are a variety of demonstratives which contain built-in sortals: 'he' for 'that male', 'she' for 'that female',<sup>46</sup> etc., and there are demonstrative phrases built from a pure demonstrative and a common noun phrase: 'that man drinking a martini', etc. Words and phrases which have demonstrative use may have other uses as well, for example, as bound variable or pronouns of laziness (anaphoric use).

I accept, tentatively and cautiously, the Fregean theory of demonstrations according to which:

- (1) A demonstration is a way of presenting an individual.
- (2) A given demonstration in certain counterfactual circumstances, would have demonstrated (i.e., presented) an individual other than the individual actually demonstrated.
- (3) A demonstration which fails to demonstrate any individual might have demonstrated one, and a demonstration which demonstrates an individual might have demonstrated no individual at all.

So far we have asserted that it is not an essential property of a given demonstration (according to the Fregean theory) that it demonstrate a given individual, or indeed, that it demonstrate any individual at all. It is this feature of demonstrations: that demonstrations which in fact demonstrate the same individual might have demonstrated distinct individuals, which provides a solution to the demonstrative version of Frege's problem (why is an utterance of 'that [Hes] = that [Phos]' informative?) analogous to Frege's own solution to the definite description version. There is some theoretical latitude as to how we should regard such other features of a demonstration as its place, time, and agent. Just to fix ideas, let us regard all these features as accidental. (It may be helpful to think of demonstrations as *types* and particular performances of them as their *tokens*). Then,

- (4) A given demonstration might have been mounted by someone other than its actual agent, and might be repeated in the same or a different place.

Although we are not now regarding the actual place and time of a demonstration as essential to it, it does seem to me to be essential to a demonstration that it present its demonstrata from some perspective, that is, as the individual that looks thusly *from here now*. On the other hand, it does not seem to me to be essential to a demonstration that it be mounted by any agent at all.<sup>47</sup>

<sup>47</sup>If the current speculations are accepted, then in the original discussion of Pat and Mike the emphasis on the counterfactual situation in which the same agent was doing the pointing was misguided and that feature of counterfactual situations is irrelevant. It is the agent of course who focuses your attention on the relevant local individual. But that needn't be done *by* anyone; we might have a convention that whoever is appearing on the demonstration platform is the demonstratum, or the speaker might take advantage of a natural demonstration of opportunity: an explosion or a shooting star.

<sup>46</sup>'Male' and 'female' are here used in the grammatical sense of gender, not the biological sense.

We now have a kind of standard form for demonstrations:

The individual that has appearance *A* from here now

where an appearance is something like a picture with a little arrow pointing to the relevant subject. Trying to put it into words, a particular demonstration might come out like:

The brightest heavenly body now visible from here.

In this example we see the importance of perspective. The same demonstration, differently located, may present a different demonstratum (a twin, for example).

If we set a demonstration,  $\delta$ , in a context, *c*, we determine the relevant perspective (i.e., the values of 'here' and 'now'). We also determine the demonstratum, if there is one—if, that is, in the circumstances of the context there is an individual that appears that way from the place and time of the context.<sup>48</sup> In setting  $\delta$  and *c* we determine more than just the demonstratum in the possible world of the context. By fixing the perspective, we determine for each possible circumstance what, if anything, would appear like that from that perspective. This is to say, we determine a *content*. This content will not, in general, be fixed (like that determined by a rigid designator). Although it was Venus that appeared a certain way from a certain location in ancient Greece, it might have been Mars. Under certain counterfactual conditions, it *would* have been Mars that appeared just that way from just that location. Set in a different context,  $\delta$ , may determine a quite different content or no content at all. When I look at myself in the mirror each morning I know that I didn't look like that ten years ago—and I suspect that nobody did.

The preceding excursion into a more detailed Fregean theory of demonstrations was simply in order to establish the following structural features of demonstrations:

1. A demonstration, when set in a context (i.e., an *occurrence* of a demonstration), determines a content.

<sup>48</sup>Since, as remarked earlier, the speaker and different members of the audience generally have different perspectives on the demonstration, it may appear slightly different to each of them. Thus each may take a slightly different demonstration to have been performed. Insofar as the agent and audience of a given context can differ in location, the location of a context is the location of the agent. Therefore the demonstratum of a given demonstration set in a given context will be the individual, if any, thereby demonstrated from the speaker's point of view.

2. It is not required that an occurrence of a demonstration have a fixed content.

In view of these features, we can associate with each demonstration a *character* which represents the 'meaning' or manner of presentation of the demonstration. We have now brought the semantics of demonstrations and descriptions into isomorphism.<sup>49</sup> Thus, I regard my 'dthat' operator as representing the general case of a demonstrative. Demonstratives are incomplete expressions which must be completed by a demonstratum (type). A complete sentence (type) will include an associated demonstration (type) for each of its demonstratives. Thus each demonstrative, *d*, will be accompanied by a demonstration,  $\delta$ , thus:

$d[\delta]$

The character of a *complete* demonstrative is given by the semantical rule:

In any context *c*,  $d[\delta]$  is a directly referential term that designates the demonstratum, if any, of  $\delta$  in *c*, and that otherwise designates nothing.

Obvious adjustments are to be made to take into account any common noun phrase which accompanies or is built-in to the demonstrative.

Since no immediately relevant structural differences have appeared between demonstrations and descriptions, I regard the treatment of the 'dthat' operator in the formal logic LD as accounting for the general case. It would be a simple matter to add to the syntax a category of 'nonlogical demonstration constants'. (Note that the indexicals of LD are all logical signs in the sense that their meaning [character] is not given by the structure but by the evaluation rules.)

## XVI. Alternative Treatments of Demonstrations

The foregoing development of the Fregean theory of demonstrations is not inevitable. Michael Bennett has proposed that only places be demonstrata and that we require an explicit or implicit common noun phrase to accompany the demonstrative, so that:

<sup>49</sup>We should not, of course, forget the many disanalogies noted earlier nor fail to note that though a description is associated with a particular character by linguistic convention, a demonstration is associated with its character by nature.



that [pointing at a person]

becomes

dthat [the person who is there [pointing at a place]].

My findings do not include the claim that the—or better, a—Fregean theory of demonstrations is correct. I can provide an alternative account for those who regard demonstrations as nonrepeatable nonseparable features of contexts. The conception now under consideration is that in certain contexts the agent is demonstrating something, or more than one thing, and in others not. Thus just as we can speak of agent, time, place, and possible world history as features of a context, we may also speak of first demonstratum, second demonstratum, . . . (some of which may be null) as features of a context. We then attach subscripts to our demonstratives and regard the  $n$ -th demonstrative, when set in a context, as rigid designator of the  $n$ -th demonstratum of the context. Such a rule associates a character with each demonstrative. In providing no role for demonstrations as separable ‘manners of presentation’ this theory eliminates the interesting distinction between demonstratives and other indexicals. We might call it the *Indexical theory of demonstratives*. (Of course every reasonable theory of demonstratives treats them as indexicals of some kind. I regard my own theory of indexicals in general, and the nondemonstrative indexicals in particular, as essentially uncontroversial. Therefore I reserve *Indexical theory of demonstratives* for the controversial alternative to the Fregean theory of demonstrations—the Fregean theory of demonstratives having been refuted.)

Let us call my theory as based on the Fregean theory of demonstrations the *Corrected Fregean theory of demonstratives*. The Fregean theory of demonstrations may be extravagant, but compared with its riches, the indexical theory is a mean thing. From a logical point of view, the riches of the Corrected Fregean theory of demonstratives are already available in connection with the demonstrative ‘dthat’ and its descriptive pseudodemonstrations, so a decision to enlarge the language of LD with additional demonstratives whose semantics are in accord with the Indexical theory need not be too greatly lamented.

If we consider Frege’s problem, we have the two formulations:

that [Hes] = that [Phos]

and

that<sub>1</sub> = that<sub>2</sub>

Both provide their sentence with an informative character. But the Fregean idea that that very demonstration might have picked out a different demonstratum seems to me to capture more of the epistemological situation than the Indexicalist’s idea that in some contexts the first and second demonstrata differ.

The Corrected Fregean theory, by incorporating demonstration types in its sentence types, accounts for more differences in informativeness as differences in meaning (character). It thereby provides a nice Frege-type solution to many Frege-type problems. But it can only forestall the resort to directly epistemological issues, it cannot hold them in abeyance indefinitely. Therefore I turn to epistemological remarks.

## XVII. Epistemological Remarks<sup>50</sup>

How do content and character serve as objects of thought? Let us state, once again, Frege’s problem

(FP) How can (an occurrence of) ‘ $\alpha = \beta$ ’ (in a given context), if true, differ in cognitive significance from (an occurrence of) ‘ $\alpha = \alpha$ ’ (in the same context)?

In (FP)  $\alpha, \beta$  are arbitrary singular terms. (In future formulations, I will omit the parentheticals as understood.) When  $\alpha$  and  $\beta$  are demonstrative free, Frege explained the difference in terms of his notion of sense. A notion which, his writings generally suggest, should be identified with our *content*. But it is clear that Frege’s problem can be reinstated in a form in which resort to contents will not explain differences in ‘cognitive significance’. We need only ask,

(FPD) How can ‘dthat[ $\alpha$ ] = dthat[ $\beta$ ]’ if true, differ in cognitive significance from ‘dthat[ $\alpha$ ] = dthat[ $\alpha$ ]’?

Since, as we shall show, for any term  $\gamma$ ,

‘ $\gamma = \text{dthat}[\gamma]$ ’ is analytic

the sentence pair in (FP) will differ in cognitive significance if and only if the sentence pair in (FPD) differ similarly. [There are a few assumptions built in here, but they are O.K.] Note, however, that the *content* of ‘dthat[ $\alpha$ ]’ and the *content* of ‘dthat[ $\beta$ ]’ are the same whenever ‘ $\alpha = \beta$ ’

<sup>50</sup>This section has benefited from the opportunity to read, and discuss with him, John Perry’s paper “Frege on Demonstratives.”

is true. Thus the difference in cognitive significance between the sentence pair in (FPD) cannot be accounted for in terms of content.

If Frege's solution to (FP) was correct, then  $\alpha$  and  $\beta$  have different contents. From this it follows that  $\ulcorner \text{dthat}[\alpha] \urcorner$  and  $\ulcorner \text{dthat}[\beta] \urcorner$  have different characters. [It doesn't really, because of the identification of contents with intensions, but let it pass.] Is character, then, the object of thought?

If you and I both say to ourselves,

(B) "I am getting bored"

have we thought the same thing? We could not have, because what you thought was true while what I thought was false.

What we must do is disentangle two epistemological notions: the *objects of thought* (what Frege called "Thoughts") and the *cognitive significance of an object of thought*. As has been noted above, a character may be likened to a manner of presentation of a content. This suggests that we identify objects of thought with contents and the cognitive significance of such objects with characters.

**E. Principle 1** *Objects of thought (Thoughts) = Contents*

**E. Principle 2** *Cognitive significance of a Thought = Character*

According to this view, the thoughts associated with  $\ulcorner \text{dthat}[\alpha] = \text{dthat}[\beta] \urcorner$  and  $\ulcorner \text{dthat}[\alpha] = \text{dthat}[\alpha] \urcorner$  are the same, but the thought (not the denotation, mind you, but the *thought*) is *presented* differently.

It is important to see that we have not *simply* generalized Frege's theory, providing a higher order Fregean sense for each name of a regular Fregean sense.<sup>51</sup> In Frege's theory, a given manner of presentation presents the same object to all mankind.<sup>52</sup> But for us, a given manner of presentation—a character—what we both said to ourselves when we both said (B)—will, in general, present different objects (of thought) to different persons (and even different Thoughts to the same person at different times).

<sup>51</sup> According to Church, such higher order Fregean senses are already called for by Frege's theory.

<sup>52</sup> See his remarks in "On Sense and Nominatum" regarding the "common treasure of thoughts which is transmitted from generation to generation" and remarks there and in "The Thought" in connection with tensed sentences, that "Only a sentence supplemented by a time-indication and complete in every respect expresses a thought."

How then can we claim that we have captured the idea of cognitive significance? To break the link between cognitive significance and universal Fregean senses and at the same time forge the link between cognitive significance and character we must come to see the *context-sensitivity* (dare I call it ego-orientation?) of cognitive states.

Let us try a Putnam-like experiment. We raise two identical twins, Castor and Pollux, under qualitatively identical conditions, qualitatively identical stimuli, etc. If necessary, we may monitor their brain states and make small corrections in their brain structures if they begin drifting apart. They respond to all cognitive stimuli in identical fashion.<sup>53</sup> Have we not been successful in achieving the same cognitive (i.e., psychological) state? Of course we have, what more could one ask! But wait, they believe different things. Each sincerely says,

My brother was born before I was

and the beliefs they thereby express conflict. In this, Castor speaks the truth, while Pollux speaks falsely. This does not reflect on the identity of their cognitive states, for, as Putnam has emphasized, circumstances alone do not determine extension (here, the truth-value) from cognitive state. Insofar as distinct persons can be in the same cognitive state, Castor and Pollux are.

**E. Corollary 1** *It is an almost inevitable consequence of the fact that two persons are in the same cognitive state, that they will disagree in their attitudes toward some object of thought.*

The corollary applies equally well to the same person at different times, and to the same person at the same time in different circumstances.<sup>54</sup> In general, the corollary applies to any individuals  $x, y$  in different contexts.

My aim was to argue that the cognitive significance of a word or phrase was to be identified with its character, the way the content is presented to us. In discussing the twins, I tried to show that persons

<sup>53</sup> Perhaps it should be mentioned here, to forestall an objection, that neither uses a proper name for the other or for himself—only 'my brother' and 'I'—and that raising them required a lot of environmental work to maintain the necessary symmetries, or, alternatively, a lot of work with the brain state machine. If proper names are present, and each uses a different name for himself (or, for the other), they will never achieve the same *total* cognitive state since one will sincerely say, "I am Castor" and the other will not. They may still achieve the same cognitive state in its relevant part.

<sup>54</sup> The corollary would also apply to the same person at the same time in the same circumstances but in different places, if such could be.

could be in the same total cognitive state and still, as we would say, believe different things. This doesn't prove that the cognitive content of, say, a single sentence or even a word is to be identified with its character, but it strongly suggests it.

Let me try a different line of argument. We agree that a given content may be presented under various characters and that consequently we may hold a propositional attitude toward a given content under one character but not under another. (For example, on March 27 of this year, having lost track of the date, I may continue to hope to be finished by this March 26, without hoping to be finished by yesterday.) Now instead of arguing that character is what we would ordinarily call cognitive significance, let me just ask why we should be interested in the character under which we hold our various attitudes. Why should we be interested in that special kind of significance that is sensitive to the use of indexicals; 'I', 'here', 'now', 'that', and the like? John Perry, in his stimulating and insightful paper "Frege on Demonstratives" asks and answers this question. [Perry uses 'thought' where I would use 'object of thought' or 'content', he uses 'apprehend' for 'believe' but *note that other psychological verbs would yield analogous cases*. I have taken a few liberties in substituting my own terminology for Perry's and have added the emphasis.]

Why should we care under what character someone apprehends a thought, so long as he does? I can only sketch the barest suggestion of an answer here. *We use the manner of presentation, the character, to individuate psychological states, in explaining and predicting action*. It is the manner of presentation, the character and not the thought apprehended, that is tied to human action. When you and I have beliefs under the common character of 'A bear is about to attack me', we behave similarly. We both roll up in a ball and try to be as still as possible. Different thoughts apprehended, same character, same behavior. When you and I both apprehend that I am about to be attacked by a bear, we behave differently. I roll up in a ball, you run to get help. Same thought apprehended, different characters, different behaviors.<sup>55</sup>

Perry's examples can be easily multiplied. My hope to be finished by a certain time is sensitive to how the content corresponding to the

<sup>55</sup> John Perry, "Frege on Demonstratives," p. 494.

time is presented, as 'yesterday' or as 'this March 26'. If I see, reflected in a window, the image of a man whose pants appear to be on fire, my behavior is sensitive to whether I think, 'His pants are on fire' or 'My pants are on fire', though the object of thought may be the same.

So long as Frege confined his attention to indexical free expressions, and given his theory of proper names, it is not surprising that he did not distinguish objects of thought (content) from cognitive significance (character), for that is the realm of *fixed* character and thus, as already remarked, there is a natural identification of character with content. Frege does, however, discuss indexicals in two places. The first passage, in which he discusses 'yesterday' and 'today' I have already discussed. Everything he says there is essentially correct. (He does not go far enough.) The second passage has provoked few endorsements and much skepticism. It too, I believe, is susceptible of an interpretation which makes it essentially correct. I quote it in full.

Now everyone is presented to himself in a particular and primitive way, in which he is presented to no one else. So, when Dr. Lauben thinks that he has been wounded, he will probably take as a basis this primitive way in which he is presented to himself. And only Dr. Lauben himself can grasp thoughts determined in this way. But now he may want to communicate with others. He cannot communicate a thought which he alone can grasp. Therefore, if he now says 'I have been wounded', he must use the 'I' in a sense that can be grasped by others, perhaps in the sense of 'he who is speaking to you at this moment', by doing which he makes the associated conditions of his utterance serve for the expression of his thought.<sup>56</sup>

What is the particular and primitive way in which Dr. Lauben is presented to himself? What cognitive content presents Dr. Lauben to himself, but presents him to nobody else? Thoughts determined this way can be grasped by Dr. Lauben, but no one else can grasp *that* thought determined in *that* way. The answer, I believe, is, simply, that Dr. Lauben is presented to himself under the character of 'I'.

A sloppy thinker might succumb to the temptation to slide from an acknowledgement of the privileged *perspective* we each have on ourselves—only I can refer to me as 'I'—to the conclusions: first, that

<sup>56</sup> Gottlob Frege, "The Thought: A Logical Inquiry," p. 298.

this perspective necessarily yields a privileged *picture* of what is seen (referred to), and second, that this picture is what is intended when one makes use of the privileged perspective (by saying 'I'). These conclusions, even if correct, are not forced upon us. The character of 'I' provides the acknowledged privileged perspective, whereas the analysis of the content of particular occurrences of 'I' provides for (and needs) no privileged pictures. There may be metaphysical, epistemological, or ethical reasons why I (so conceived) am especially *important* to myself. (Compare: why *now* is an especially important time to me. It too is presented in a particular and primitive way, and this moment cannot be presented at any other time in the same way.)<sup>57</sup> But the phenomenon noted by Frege—that everyone is presented to himself in a particular and primitive way—can be fully accounted for using only our semantical theory.

Furthermore, regarding the first conclusion, I sincerely doubt that there is, for each of us on each occasion of the use of 'I', a particular, primitive, and incommunicable Fregean self-concept which we tacitly express to ourselves. And regarding the second conclusion: even if Castor were sufficiently narcissistic to associate such self-concepts with his every use of 'I', his twin, Pollux, whose mental life is qualitatively identical with Castor's, would associate the *same* self-concept with *his* every (matching) use of 'I'.<sup>58</sup> The second conclusion would lead to the absurd result that when Castor and Pollux each say 'I', they do not thereby distinguish themselves from one another. (An even more astonishing result is possible. Suppose that due to a bit of self-deception the self-concept held in common by Castor and Pollux fits neither of them. The second conclusion then leads irresistibly to the possibility that when Castor and Pollux each say 'I' they each refer to a third party!)

The perceptive reader will have noticed that the conclusions of the sloppy thinker regarding the pure indexical 'I' are not unlike those of the Fregean regarding true demonstratives. The sloppy thinker has adopted a *demonstrative theory of indexicals*: 'I' is synonymous with 'this person' [along with an appropriate *subjective* demonstration], 'now' with 'this time', 'here' with 'this place' [each associated with some demonstration], etc. Like the Fregean, the sloppy thinker errs in believing that the

sense of the demonstration is the sense of the indexical, but the sloppy thinker commits an additional error in believing that such senses are in any way necessarily associated with uses of pure indexicals. The slide from privileged perspective to privileged picture is the sloppy thinker's original sin. Only one who is located in the exact center of the Sahara Desert is entitled to refer to that place as 'here', but aside from that, the place may present no distinguishing features.<sup>59</sup>

The sloppy thinker's conclusions may have another source. Failure to distinguish between the cognitive significance of a thought and the thought itself seems to have led some to believe that the elements of an object of thought must each be directly accessible to the mind. From this it follows that if a singular proposition is an object of thought, the thinker must somehow be immediately acquainted with each of the individuals involved. But, as we have seen, the situation is rather different from this. Singular propositions may be presented to us under characters which neither imply nor presuppose any special form of acquaintance with the individuals of the singular propositions. The psychological states, perhaps even the epistemological situations, of Castor and Pollux are alike, yet they assert distinct singular propositions when they each say 'My brother was born before me'. Had they lived at different times they might still have been situated alike epistemologically

<sup>59</sup> So far, we have limited our attention to the first three sentences of the quotation from Frege. How are we to account for the second part of Frege's remarks?

Suppose Dr. Lauben wants to communicate his thought without disturbing its cognitive content. (Think of trying to tell a color-blind person that the green light should be replaced. You would have to find another way of communicating what you wanted to get across.) He can't communicate *that* thought with *that* significance, so, he himself would have to attach a nonstandard significance to 'I'. Here is a suggestion. He points at his auditor and uses the demonstrative 'you'. If we neglect fine differences in perspective, the demonstration will have the same character for all present and it certainly will have the same demonstratum for all present, therefore the demonstrative will have the same *character and content* for all present. The indexical 'now' will certainly have the same character and content for all present. Thus 'the person who is speaking to you [points] now' will have a common character and content for all those present. Unfortunately the content is not that of 'I' as Dr. Lauben standardly uses it. He needs a demonstrative like 'dthat' to convert the description to a term with a fixed content. He chooses the demonstrative 'he', with a relative clause construction to make clear his intention. Now, if Dr. Lauben uses 'I' with the nonstandard meaning usually attached to 'he who is speaking to you [points] now' he will have found a way to communicate his original thought in a form whose cognitive significance is common to all. Very clever, Dr. Lauben.

[Perhaps it is poor pedagogy to join this fanciful interpretation of the second part of the passage with the serious interpretation of the first part.]

<sup>57</sup> At other times, earlier and later, we can know it only externally, by description as it were. But now we are directly acquainted with it. (I believe I owe this point to John Perry.)

<sup>58</sup> Unless, of course, the self-concept involved a bit of direct reference. In which case (when direct reference is admitted) there seems no need for the whole theory of Fregean self-concepts. Unless, of course, direct reference is limited to items of direct acquaintance, of which more below.

while asserting distinct singular propositions in saying 'It is quiet here now'. A kidnapped heiress, locked in the trunk of a car, knowing neither the time nor where she is, may think 'It is quiet here now' and the indexicals will remain directly referential.<sup>60</sup>

**E. Corollary 2** *Ignorance of the referent does not defeat the directly referential character of indexicals.*

From this it follows that a special form of knowledge of an object is neither required nor presupposed in order that a person may entertain as object of thought a singular proposition involving that object.

There is nothing inaccessible to the mind about the semantics of direct reference, even when the reference is to that which we know only by description. What allows us to take various propositional attitudes towards singular propositions is not the form of our acquaintance with the objects but is rather our ability to manipulate the conceptual apparatus of direct reference.<sup>61</sup>

The foregoing remarks are aimed at refuting *Direct Acquaintance Theories of direct reference*. According to such theories, the question whether an utterance expresses a singular proposition turns, in the first instance, on the speaker's *knowledge of the referent* rather than on the *form of the reference*. If the speaker lacks the appropriate form of acquaintance with the referent, the utterance cannot express a singular proposition, and any apparently directly referring expressions used must be abbreviations or disguises for something like Fregean descriptions. Perhaps the Direct Acquaintance theorist thought that only a theory like his could permit singular propositions while still providing a solution for Frege's problem. If we could *directly* refer to a given object in nonequivalent ways (e.g., as 'dthat[Hes]' and 'dthat[Phos]'), we could not—so he thought—explain the difference in cognitive significance between the appropriate instances of ' $\alpha = \alpha$ ' and ' $\alpha = \beta$ '. Hence, the objects susceptible to direct reference must not permit such reference in inequivalent ways. These objects must, in a certain sense, be wholly local and completely given so that for any two *directly* coreferential terms

<sup>60</sup> Can the heiress plead that she could not have believed a singular proposition involving the place *p* since when thinking 'here' she didn't *know* she was at *p*, that she was, in fact, unacquainted with the place *p*? No! Ignorance of the referent is no excuse.

<sup>61</sup> This makes it sound as if an exact and conscious mastery of semantics is prerequisite to having a singular proposition as object of thought. I will try to find a better way to express the point in a succeeding draft.

$\alpha$  and  $\beta$ , ' $\alpha = \beta$ ' will be uninformative to anyone appropriately situated, epistemologically, to be able to use these terms.<sup>62</sup> I hope that my discussion of the two kinds of meaning—content and character—will have shown the Direct Acquaintance Theorist that his views are not the inevitable consequence of the admission of directly referential terms. From the point of view of a lover of direct reference this is good, since the Direct Acquaintance theorist admits direct reference in a portion of language so narrow that it is used only by philosophers.<sup>63</sup>

I have said nothing to dispute the epistemology of the Direct Acquaintance theorist, nothing to deny that there exists his special kind of object with which one can have his special kind of acquaintance. I have only denied the relevance of these epistemological claims to the semantics of direct reference. If we sweep aside metaphysical and epistemological pseudo-explanations of what are essentially semantical phenomena, the result can only be healthy for all three disciplines.

Before going on to further examples of the tendency to confuse metaphysical and epistemological matters with phenomena of the semantics of direct reference, I want to briefly raise the problem of *cognitive dynamics*. Suppose that yesterday you said, and believed it, "It is a nice day today." What does it mean to say, today, that you have retained *that* belief? It seems unsatisfactory to just believe the same content under any old character—where is the *retention*?<sup>64</sup> You *can't* believe

<sup>62</sup> For some consequences of this view with regard to the interpretation of demonstratives see "Bob and Carol and Ted and Alice," appendix VII.

<sup>63</sup> There is an obvious connection between the fix in which the Direct Acquaintance Theorist finds himself, and *Kripke's problem*: how can ' $\alpha = \beta$ ' be informative if  $\alpha$  and  $\beta$  differ in neither denotation nor sense (nor, as I shall suggest is the case for proper names, character)?

<sup>64</sup> The sort of case I have in mind is this. I first think, "His pants are on fire." I later realize, "I am he" and thus come to think "My pants are on fire." Still later, I decide that I was wrong in thinking "I am he" and conclude "His pants were on fire." If, in fact, I *am* he, have I *retained* my belief that my pants are on fire simply because I believe the same content, though under a different character? (I also deny that content under the former, but for change of tense, character.) When I first thought "My pants are on fire," a certain singular proposition, call it 'Eek', was the object of thought. At the later stage, both Eek and its negation are believed by me. In this sense, I still believe what I believed before, namely Eek. But this does not capture my sense of *retaining a belief*: a sense that I associate with saying that some people have a very rigid cognitive structure whereas others are very flexible. It is tempting to say that cognitive dynamics is concerned not with retention and change in what is believed, but with retention and change in the characters under which our beliefs are held. I think that this is basically correct. But it is not obvious to me what relation between a character under which a belief is held at one time and the set of characters under which beliefs are held at a later

that content under the same character. Is there some obvious standard adjustment to make to the character, for example, replacing *today* with *yesterday*? If so, then a person like Rip van Winkle, who loses track of time, can't retain any such beliefs. This seems strange. Can we only *retain* beliefs presented under a fixed character? This issue has obvious and important connections with Lauben's problem in trying to communicate the thought he expresses with 'I have been wounded'. Under what character must his auditor believe Lauben's thought in order for Lauben's communication to have been successful? It is important to note that if Lauben said 'I am wounded' in the usual meaning of 'I', there is no one else who can report what he said, using *indirect* discourse, and convey the cognitive significance (to Lauben) of what he said. This is connected with points made in section VIII, and has interesting consequences for the inevitability of so-called *de re* constructions in indirect discourse languages which contain indexicals. (I use 'indirect discourse' as a general term for the analogous form of all psychological verbs.)

A prime example of the confusion of direct reference phenomena with metaphysical and epistemological ideas was first vigorously called to our attention by Saul Kripke in *Naming and Necessity*. I wish to parallel his remarks disconnecting the *a priori* and the *necessary*.

The form of a *prioricity* that I will discuss is that of logical truth (in the logic of demonstratives). We saw very early that a truth of the logic of demonstratives, like "I am here now" need not be necessary. There are many such cases of logical truths which are not necessary. If  $\alpha$  is any singular term, then

$$\alpha = \text{dthat}[\alpha]$$

is a logical truth. But

$$\Box(\alpha = \text{dthat}[\alpha])$$

is generally false. We can, of course, also easily produce the opposite effect.

---

time would constitute retaining the original belief. Where indexicals are involved, for the reasons given below, we cannot simply require that the very same character still appear at the later time. Thus the problem of cognitive dynamics can be put like this: what does it mean to say of an individual who at one time sincerely asserted a sentence containing indexicals that at some later time he has (or has not) *changed his mind* with respect to his assertion? What sentence or sentences must he be willing to assert at the later time?

$$\Box(\text{dthat}[\alpha] = \text{dthat}[\beta])$$

may be true, although

$$\text{dthat}[\alpha] = \text{dthat}[\beta]$$

is not logically true, and is even logically equivalent to the contingency,

$$\alpha = \beta$$

(I call  $\phi$  and  $\psi$  logically equivalent when ' $\phi \leftrightarrow \psi$ ' is logically true.) These cases are reminiscent of Kripke's case of the terms, 'one meter' and 'the length of bar  $x$ '. But where Kripke focuses on the special epistemological situation of one who is present at the dubbing, the descriptive meaning associated with our directly referential term  $\text{dthat}[\alpha]$  is carried in the semantics of the language.<sup>65</sup>

How can something be both logically true, and thus *certain*, and *contingent* at the same time? In the case of indexicals the answer is easy to see.

**E. Corollary 3** *The bearers of logical truth and of contingency are different entities. It is the character (or, the sentence, if you prefer) that is logically true, producing a true content in every context. But it is the content (the proposition, if you will) that is contingent or necessary.*

As can readily be seen, the modal logic of demonstratives is a rich and interesting thing.

---

<sup>65</sup> A case of a seemingly different kind is that of the logical equivalence between an arbitrary sentence  $\phi$  and the result of prefixing either or both of the indexical operators, 'it is actually the case that' (symbolized ' $A$ ') and 'it is now the case that' (symbolized ' $N$ '). The biconditional ' $\phi \leftrightarrow AN\phi$ ' is logically true, but prefixing either ' $\Box$ ' or its temporal counterpart can lead to falsehood. (This case was adverted to in footnote 28.) It is interesting to note, in this case, that the parallel between modal and temporal modifications of sentences carries over to indexicals. The foregoing claims are verified by the formal system (sections XVIII and XIX, see especially Remark 3). Note that the formal system is constructed in accordance with Carnap's proposal that the intension of an expression be that function which assigns to each circumstance, the extension of the expression with respect to that circumstance. This has commonly been thought to insure that logically equivalent expressions have the same intension (Church's Alternative 2 among principles of individuation for the notion of sense) and that logically true sentences express the (unique) necessary proposition. Homework Problem: What went wrong here?

It is easy to be taken in by the effortless (but fallacious) move from certainty (logical truth) to necessity. In his important article "Three Grades of Modal Involvement,"<sup>66</sup> Quine expresses his scepticism of the first grade of modal involvement: the sentence predicate and all it stands for, and his distaste for the second grade of modal involvement: disguising the predicate as an operator 'It is necessary that'. But he suggests that no new metaphysical undesirables are admitted until the third grade of modal involvement: quantification across the necessity operator into an open sentence.

I must protest. That first step let in some metaphysical undesirables, falsehoods. All logical truths are analytic, but they can go false when you back them up to '□'.

One other notorious example of a logical truth which is not necessary,

I exist.

One can quickly verify that in every context, this character yields a true proposition—but rarely a necessary one. It seems likely to me that it was a conflict between the feelings of contingency and of certainty associated with this sentence that has led to such painstaking examination of its 'proofs'. It is just a truth of logic!

Dana Scott has remedied one lacuna in this analysis. What of the premise

I think

and the connective

Therefore ?

His discovery was that the premise is incomplete, and that the last five words

up the logic of demonstratives

had been lost in an early manuscript version.<sup>67</sup>

<sup>66</sup> *Proceedings of the XI International Congress of Philosophy* 14, 65–81; reprinted in W. V. Quine, *The Ways of Paradox* (New York: Random House, 1966).

<sup>67</sup> Again, it is probably a pedagogical mistake to mix this playful paragraph with the preceding serious one.

## XVIII. The Formal System

Just to be sure we have not overlooked anything, here is a machine against which we can test our intuitions.

### The Language LD

The *Language LD* is based on first-order predicate logic with identity and descriptions. We deviate slightly from standard formulations in using two sorts of variables, one sort for positions and a second for individuals other than positions (hereafter called simply 'individuals').

### Primitive Symbols

#### Primitive Symbols for Two Sorted Predicate Logic

0. Punctuation: (, ), [, ]
1. Variables:
  - (i) An infinite set of individual variables:  $\mathcal{V}_i$
  - (ii) An infinite set of position variables:  $\mathcal{V}_p$
2. Predicates:
  - (i) An infinite number of  $m$ - $n$ -place predicates, for all natural numbers  $m, n$ .
  - (ii) The 1-0-place predicate: Exist
  - (iii) The 1-1-place predicate: Located
3. Functors:
  - (i) An infinite number of  $m$ - $n$ -place  $i$ -functors (functors which form terms denoting individuals)
  - (ii) An infinite number of  $m$ - $n$ -place  $p$ -functors (functors which form terms denoting positions)
4. Sentential Connectives:  $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$
5. Quantifiers:  $\forall, \exists$
6. Definite Description Operator: the
7. Identity: =

## Primitive Symbols for Modal and Tense Logic

8. Modal Operators:  $\Box, \Diamond$
9. Tense Operators:  
 $F$  (it will be the case that)  
 $P$  (it has been the case that)  
 $G$  (one day ago, it was the case that)

## Primitive Symbols for the Logic of Demonstratives

10. Three 1-place sentential operators:  
 $N$  (it is now the case that)  
 $A$  (it is actually the case that)  
 $Y$  (yesterday, it was the case that)
11. A 1-place functor:  $d$ that
12. An individual constant (0-0-place  $i$ -functor):  $I$
13. A position constant (0-0-place  $p$ -functor): Here

## Well-formed Expressions

The *well-formed expressions* are of three kinds: formulas, position terms ( $p$ -terms), and individual terms ( $i$ -terms).

1. (i) If  $\alpha \in \mathcal{V}_i$ , then  $\alpha$  is an  $i$ -term  
 (ii) If  $\alpha \in \mathcal{V}_p$ , then  $\alpha$  is a  $p$ -term
2. If  $\pi$  is an  $m$ - $n$ -place predicate,  $\alpha_1, \dots, \alpha_m$  are  $i$ -terms, and  $\beta_1, \dots, \beta_n$  are  $p$ -terms, then  $\pi\alpha_1 \dots \alpha_m \beta_1 \dots \beta_n$  is a formula
3. (i) If  $\eta$  is an  $m$ - $n$ -place  $i$ -functor,  $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_n$  are as in 2., then  
 $\eta\alpha_1 \dots \alpha_m \beta_1 \dots \beta_n$  is an  $i$ -term  
 (ii) If  $\eta$  is an  $m$ - $n$ -place  $p$ -functor,  $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_n$  are as in 2., then  
 $\eta\alpha_1 \dots \alpha_m \beta_1 \dots \beta_n$  is a  $p$ -term
4. If  $\phi, \psi$  are formulas, then  $(\phi \wedge \psi), (\phi \vee \psi), \neg\phi, (\phi \rightarrow \psi), (\phi \leftrightarrow \psi)$  are formulas
5. If  $\phi$  is a formula and  $\alpha \in \mathcal{V}_i \cup \mathcal{V}_p$ , then  $\forall\alpha\phi$  and  $\exists\alpha\phi$  are formulas

6. If  $\phi$  is a formula, then  
 (i) if  $\alpha \in \mathcal{V}_i$ , then the  $\alpha\phi$  is an  $i$ -term  
 (ii) if  $\alpha \in \mathcal{V}_p$ , then the  $\alpha\phi$  is a  $p$ -term
7. If  $\alpha, \beta$  are either both  $i$ -terms or both  $p$ -terms, then  $\alpha = \beta$  is a formula
8. If  $\phi$  is a formula, then  $\Box\phi$  and  $\Diamond\phi$  are formulas
9. If  $\phi$  is a formula, then  $F\phi, P\phi$ , and  $G\phi$  are formulas
10. If  $\phi$  is a formula, then  $N\phi, A\phi$ , and  $Y\phi$  are formulas
11. (i) If  $\alpha$  is an  $i$ -term, then  $d$ that $[\alpha]$  is an  $i$ -term  
 (ii) If  $\alpha$  is a  $p$ -term, then  $d$ that $[\alpha]$  is a  $p$ -term

## Semantics for LD

## LD Structures

**Definition:**  $\mathfrak{A}$  is an LD structure iff there are  $\mathcal{C}, \mathcal{W}, \mathcal{U}, \mathcal{P}, \mathcal{T}$ , and  $\mathcal{I}$  such that:

1.  $\mathfrak{A} = \langle \mathcal{C}, \mathcal{W}, \mathcal{U}, \mathcal{P}, \mathcal{T}, \mathcal{I} \rangle$
2.  $\mathcal{C}$  is a nonempty set (the set of contexts, see 10 below)
3. If  $c \in \mathcal{C}$ , then  
 (i)  $c_A \in \mathcal{U}$  (the *agent* of  $c$ )  
 (ii)  $c_T \in \mathcal{T}$  (the *time* of  $c$ )  
 (iii)  $c_P \in \mathcal{P}$  (the *position* of  $c$ )  
 (iv)  $c_W \in \mathcal{W}$  (the *world* of  $c$ )
4.  $\mathcal{W}$  is a nonempty set (the set of *worlds*)
5.  $\mathcal{U}$  is a nonempty set (the set of all *individuals*, see 9 below)
6.  $\mathcal{P}$  is a nonempty set (the set of *positions*, common to all worlds)
7.  $\mathcal{T}$  is the set of integers (thought of as the *times*, common to all worlds)



8.  $\mathcal{I}$  is a function which assigns to each predicate and functor an appropriate *intension* as follows:
- (i) If  $\pi$  is an  $m$ - $n$ -predicate,  $\mathcal{I}_\pi$  is a function such that for each  $t \in \mathcal{T}$  and  $w \in \mathcal{W}$ ,  $\mathcal{I}_\pi(t, w) \subseteq (\mathcal{U}^m \times \mathcal{P}^n)$
  - (ii) If  $\eta$  is an  $m$ - $n$ -place  $i$ -functor,  $\mathcal{I}_\eta$  is a function such that for each  $t \in \mathcal{T}$  and  $w \in \mathcal{W}$ ,  $\mathcal{I}_\eta(t, w) \in (\mathcal{U} \cup \{\dagger\})^{(\mathcal{U}^m \times \mathcal{P}^n)}$  (Note:  $\dagger$  is a completely alien entity, in neither  $\mathcal{U}$  nor  $\mathcal{P}$ , which represents an 'undefined' value of the function. In a normal set theory we can take  $\dagger$  to be  $\{\mathcal{U}, \mathcal{P}\}$ .)
  - (iii) If  $\eta$  is an  $m$ - $n$ -place  $p$ -functor,  $\mathcal{I}_\eta$  is a function such that for each  $t \in \mathcal{T}$  and  $w \in \mathcal{W}$ ,  $\mathcal{I}_\eta(t, w) \in (\mathcal{P} \cup \{\dagger\})^{(\mathcal{U}^m \times \mathcal{P}^n)}$
9.  $i \in \mathcal{U}$  iff  $(\exists t \in \mathcal{T})(\exists w \in \mathcal{W})(\langle i \rangle \in \mathcal{I}_{\text{Exist}}(t, w))$
10. If  $c \in \mathcal{C}$ , then  $\langle c_A, c_P \rangle \in \mathcal{I}_{\text{Located}}(c_T, c_W)$
11. If  $\langle i, p \rangle \in \mathcal{I}_{\text{Located}}(t, w)$ , then  $\langle i \rangle \in \mathcal{I}_{\text{Exist}}(t, w)$

### Truth and Denotation in a Context

We write:  $\models_{cftw}^{\mathfrak{A}} \phi$  for  $\phi$ , when taken in the context  $c$  (under the assignment  $f$  and in the structure  $\mathfrak{A}$ ), is true with respect to the time  $t$  and the world  $w$ .

We write:  $|\alpha|_{cftw}^{\mathfrak{A}}$  for The denotation of  $\alpha$ , when taken in the context  $c$  (under the assignment  $f$  and in the structure  $\mathfrak{A}$ ), with respect to the time  $t$  and the world  $w$

In general we will omit the superscript ' $\mathfrak{A}$ ', and we will assume that the structure  $\mathfrak{A}$  is  $\langle \mathcal{C}, \mathcal{W}, \mathcal{U}, \mathcal{P}, \mathcal{T}, \mathcal{I} \rangle$ .

**Definition:**  $f$  is an assignment (with respect to  $\langle \mathcal{C}, \mathcal{W}, \mathcal{U}, \mathcal{P}, \mathcal{T}, \mathcal{I} \rangle$ ) iff:

$$\exists f_1 f_2 (f_1 \in \mathcal{U}^{\mathcal{V}_i} \ \& \ f_2 \in \mathcal{P}^{\mathcal{V}_p} \ \& \ f = f_1 \cup f_2)$$

**Definition:**  $f_x^\alpha = (f \sim \{(\alpha, f(\alpha))\}) \cup \{(\alpha, x)\}$

(i.e., the assignment which is just like  $f$  except that it assigns  $x$  to  $\alpha$ )

**Definition:** For the following recursive definition, assume that  $c \in \mathcal{C}$ ,  $f$  is an assignment,  $t \in \mathcal{T}$ , and  $w \in \mathcal{W}$ :

1. If  $\alpha$  is a variable,  $|\alpha|_{cftw} = f(\alpha)$
2.  $\models_{cftw} \pi \alpha_1 \dots \alpha_m \beta_1 \dots \beta_n$  iff  $\langle |\alpha_1|_{cftw} \dots |\beta_n|_{cftw} \rangle \in \mathcal{I}_\pi(t, w)$
3. If  $\eta$  is neither 'T' nor 'Here' (see 12, 13 below), then
 
$$|\eta \alpha_1 \dots \alpha_m \beta_1 \dots \beta_n|_{cftw} = \begin{cases} \mathcal{I}_\eta(t, w)(\langle |\alpha_1|_{cftw} \dots |\beta_n|_{cftw} \rangle), \\ \text{if none of } |\alpha_j|_{cftw} \dots |\beta_k|_{cftw} \\ \text{are } \dagger; \\ \dagger, \text{ otherwise} \end{cases}$$
4. (i)  $\models_{cftw} (\phi \wedge \psi)$  iff  $\models_{cftw} \phi$  &  $\models_{cftw} \psi$   
 (ii)  $\models_{cftw} \neg \phi$  iff  $\sim \models_{cftw} \phi$   
 etc.
5. (i) If  $\alpha \in \mathcal{V}_i$ , then  $\models_{cftw} \forall \alpha \phi$  iff  $\forall i \in \mathcal{U}, \models_{cf_i^* tw} \phi$   
 (ii) If  $\alpha \in \mathcal{V}_p$ , then  $\models_{cftw} \forall \alpha \phi$  iff  $\forall p \in \mathcal{P}, \models_{cf_p^* tw} \phi$   
 (iii) Similarly for  $\exists \alpha \phi$
6. (i) If  $\alpha \in \mathcal{V}_i$ , then:
 
$$|\text{the } \alpha \phi|_{cftw} = \begin{cases} \text{the unique } i \in \mathcal{U} \text{ such that } \models_{cf_i^* tw} \phi, \text{ if} \\ \text{there is such;} \\ \dagger, \text{ otherwise} \end{cases}$$
 (ii) Similarly for  $\alpha \in \mathcal{V}_p$
7.  $\models_{cftw} \alpha = \beta$  iff  $|\alpha|_{cftw} = |\beta|_{cftw}$
8. (i)  $\models_{cftw} \Box \phi$  iff  $\forall w' \in \mathcal{W}, \models_{cftw'} \phi$   
 (ii)  $\models_{cftw} \Diamond \phi$  iff  $\exists w' \in \mathcal{W}, \models_{cftw'} \phi$
9. (i)  $\models_{cftw} F \phi$  iff  $\exists t' \in \mathcal{T}$  such that  $t' > t$  and  $\models_{cft'w} \phi$   
 (ii)  $\models_{cftw} P \phi$  iff  $\exists t' \in \mathcal{T}$  such that  $t' < t$  and  $\models_{cft'w} \phi$   
 (iii)  $\models_{cftw} G \phi$  iff  $\models_{cf(t-1)w} \phi$
10. (i)  $\models_{cftw} N \phi$  iff  $\models_{cf(c_T)w} \phi$   
 (ii)  $\models_{cftw} A \phi$  iff  $\models_{cftc_W} \phi$   
 (iii)  $\models_{cftw} Y \phi$  iff  $\models_{cf(c_T-1)w} \phi$

11.  $|dthat[\alpha]|_{cftw} = |\alpha|_{cfc_{\tau}cw}$   
 12.  $|I|_{cftw} = c_A$   
 13.  $|Here|_{cftw} = c_P$

### XIX. Remarks on the Formal System

**Remark 1:** Expressions containing demonstratives will, in general, express different concepts in different contexts. We call the concept expressed in a given context the *Content* of the expression in that context. The Content of a sentence in a context is, roughly, the proposition the sentence would express if uttered in that context. This description is not quite accurate on two counts. First, it is important to distinguish an *utterance* from a *sentence-in-a-context*. The former notion is from the theory of speech acts, the latter from semantics. Utterances take time, and utterances of distinct sentences cannot be simultaneous (i.e., in the same context). But to develop a logic of demonstratives it seems most natural to be able to evaluate several premises and a conclusion all in the same context. Thus the notion of  $\phi$  being true in  $c$  and  $\mathfrak{A}$  does not require an utterance of  $\phi$ . In particular,  $c_A$  need not be uttering  $\phi$  in  $c_W$  at  $c_T$ . Second, the truth of a proposition is not usually thought of as dependent on a time as well as a possible world. The time is thought of as fixed by the context. If  $\phi$  is a sentence, the more usual notion of the proposition expressed by  $\phi$ -in- $c$  is what is here called the Content of  $N\phi$  in  $c$ .

Where  $\Gamma$  is either a term or formula,

we write:  $\{\Gamma\}_{cf}^{\mathfrak{A}}$  for The Content of  $\Gamma$  in the context  $c$  (under the assignment  $f$  and in the structure  $\mathfrak{A}$ ).

#### Definition:

- (i) If  $\phi$  is a formula,  $\{\phi\}_{cf}^{\mathfrak{A}}$  = that function which assigns to each  $t \in T$  and  $w \in \mathcal{W}$ , Truth, if  $\models_{cftw}^{\mathfrak{A}} \phi$ , and Falsehood otherwise.  
 (ii) If  $\alpha$  is a term,  $\{\alpha\}_{cf}^{\mathfrak{A}}$  = that function which assigns to each  $t \in T$  and  $w \in \mathcal{W}$ ,  $|\alpha|_{cftw}$ .

**Remark 2:**  $\models_{cftw}^{\mathfrak{A}} \phi$  iff  $\{\phi\}_{cf}^{\mathfrak{A}}(t, w) = \text{Truth}$ . Roughly speaking, the sentence  $\phi$  taken in context  $c$  is true with respect to  $t$  and  $w$  iff the proposition expressed by  $\phi$ -in-the-context- $c$  would be true at the time  $t$  if  $w$  were the actual world. In the formal development of pages 544, 545, and 546, it was smoother to ignore the conceptual break marked by the notion of *Content in a context* and to directly define *truth in a context with respect to a possible time and world*. The important conceptual role of the notion of Content is partially indicated by the following two definitions.

**Definition:**  $\phi$  is true in the context  $c$  (in the structure  $\mathfrak{A}$ ) iff for every assignment  $f$ ,  $\{\phi\}_{cf}^{\mathfrak{A}}(c_T, c_W) = \text{Truth}$ .

**Definition:**  $\phi$  is valid in LD ( $\models \phi$ ) iff for every LD structure  $\mathfrak{A}$ , and every context  $c$  of  $\mathfrak{A}$ ,  $\phi$  is true in  $c$  (in  $\mathfrak{A}$ ).

**Remark 3:**  $\models (\alpha = dthat[\alpha])$ ;  $\models (\phi \leftrightarrow AN\phi)$ ;  $\models N(\text{Located } I, \text{Here})$ ;  $\models \text{Exist } I$ . But,  $\not\models \Box(\alpha = dthat[\alpha])$ ;  $\not\models \Box(\phi \leftrightarrow AN\phi)$ ;  $\not\models \Box N(\text{Located } I, \text{Here})$ ;  $\not\models \Box(\text{Exist } I)$ . Also,  $\not\models F(\phi \leftrightarrow AN\phi)$ .

In the converse direction (where the original validity has the form  $\Box\phi$ ) we have the usual results in view of the fact that  $\models (\Box\phi \rightarrow \phi)$ .

**Definition:** If  $\alpha_1, \dots, \alpha_n$  are all the free variables of  $\phi$  in alphabetical order then the closure of  $\phi = AN\forall\alpha_1 \dots \forall\alpha_n \phi$ .

**Definition:**  $\phi$  is closed iff  $\phi$  is equivalent (in the sense of Remark 12) to its closure.

**Remark 4:** If  $\phi$  is closed, then  $\phi$  is true in  $c$  (and  $\mathfrak{A}$ ) iff for every assignment  $f$ , time  $t$ , and world  $w$ ,  $\models_{cftw}^{\mathfrak{A}} \phi$ .

**Definition:** Where  $\Gamma$  is either a term or a formula, the Content of  $\Gamma$  in the context  $c$  (in the structure  $\mathfrak{A}$ ) is Stable iff for every assignment  $f$ ,  $\{\Gamma\}_{cf}^{\mathfrak{A}}$  is a constant function (i.e.,  $\{\Gamma\}_{cf}^{\mathfrak{A}}(t, w) = \{\Gamma\}_{cf}^{\mathfrak{A}}(t', w')$ , for all  $t, t', w, w'$  in  $\mathfrak{A}$ ).

**Remark 5:** Where  $\phi$  is a formula,  $\alpha$  is a term, and  $\beta$  is a variable, each of the following has a Stable Content in every context (in every structure):  $AN\phi$ ,  $dthat[\alpha]$ ,  $\beta$ ,  $I$ ,  $Here$ .

If we were to extend the notion of Content to apply to operators, we would see that all indexicals (including  $N$ ,  $A$ ,  $Y$ , and  $dthat$ ) have a Stable Content in every context. The same is true of the familiar logical constants although it does not hold for the modal and tense operators (not, at least, according to the foregoing development).

**Remark 6:** That aspect of the meaning of an expression which determines what its Content will be in each context, we call the *Character* of the expression. Although a lack of knowledge about the context (or perhaps about the structure) may cause one to mistake the Content of a given utterance, the Character of each well-formed expression is determined by rules of the language (such as rules 1–13 on pages 545 and 546, which are presumably known to all competent speakers. Our notation ' $\{\phi\}_{cf}^{\mathfrak{A}}$ ' for the Content of an expression gives a natural notation for the Character of an expression, namely ' $\{\phi\}$ '.

**Definition:** Where  $\Gamma$  is either a term or a formula, the *Character* of  $\Gamma$  is that function which assigns to each structure  $\mathfrak{A}$ , assignment  $f$ , and context  $c$  of  $\mathfrak{A}$ ,  $\{\Gamma\}_{cf}^{\mathfrak{A}}$ .

**Definition:** Where  $\Gamma$  is either a term or a formula, the *Character* of  $\Gamma$  is *Stable* iff for every structure  $\mathfrak{A}$ , and assignment  $f$ , the Character of  $\Gamma$  (under  $f$  in  $\mathfrak{A}$ ) is a constant function (i.e.,  $\{\Gamma\}_{cf}^{\mathfrak{A}} = \{\Gamma\}_{c'f}^{\mathfrak{A}}$ , for all  $c, c'$  in  $\mathfrak{A}$ ).

**Remark 7:** A formula or term has a Stable Character iff it has the same Content in every context (for each  $\mathfrak{A}, f$ ).

**Remark 8:** A formula or term has a Stable Character iff it contains no essential occurrence of a demonstrative.

**Remark 9:** The logic of demonstratives determines a sublogic of those formulas of LD which contain no demonstratives. These formulas (and their equivalents which contain inessential occurrences of demonstratives) are exactly the formulas with a Stable Character. The logic of demonstratives brings a new perspective even to formulas such as these.

The sublogic of LD which concerns only formulas of Stable Character is not identical with traditional logic. Even for such formulas, the familiar Principle of Necessitation (if  $\models \phi$ , then  $\models \Box\phi$ ) fails. And so does its tense logic counterpart: if  $\models \phi$ , then  $\models (\neg P\neg\phi \wedge \neg F\neg\phi \wedge \phi)$ . From the perspective of LD, validity is truth in every possible *context*. For traditional logic, validity is truth in every possible *circumstance*. Each possible context determines a possible circumstance, but it is not the case that each possible circumstance is part of a possible context. In particular, the fact that each possible context has an agent implies that any possible circumstance in which no individuals exist will not form a part of any possible context. Within LD, a possible context is represented by  $\langle \mathfrak{A}, c \rangle$  and a possible circumstance by  $\langle \mathfrak{A}, t, w \rangle$ . To any  $\langle \mathfrak{A}, c \rangle$ , there corresponds  $\langle \mathfrak{A}, c_T, c_W \rangle$ . But it is not the case that to every  $\langle \mathfrak{A}, t, w \rangle$  there exists a context  $c$  of  $\mathfrak{A}$  such that  $t = c_T$  and  $w = c_W$ . The result is that in LD such sentences as ' $\exists x$  Exist  $x$ ' and ' $\exists x \exists p$  Located  $x, p$ ' are valid, although they would not be so regarded in traditional logic. At least not in the neotraditional logic that countenances empty worlds. Using the semantical developments of pages 543–46, we can define this traditional sense of validity (for formulas which do not contain demonstratives) as follows. First note that by Remark 7, if  $\phi$  has a Stable Character,

$$\models_{cftw}^{\mathfrak{A}} \phi \text{ iff } \models_{c'ftw}^{\mathfrak{A}} \phi$$

Thus for such formulas we can define,

$$\phi \text{ is true at } t, w \text{ (in } \mathfrak{A}) \text{ iff for every assignment } f \text{ and every context } c, \models_{cftw}^{\mathfrak{A}} \phi$$

The neotraditional sense of validity is now definable as follows,

$$\models_T \phi \text{ iff for all structures } \mathfrak{A}, \text{ times } t, \text{ and worlds } w, \phi \text{ is true at } t, w \text{ (in } \mathfrak{A})$$

(Properly speaking, what I have called the neo-traditional sense of validity is the notion of validity now common for a quantified S5 modal tense logic with individual variables ranging over possible individuals and a predicate of existence.) Adding the subscript 'LD' for explicitness, we can now state some results.

- (i) If  $\phi$  contains no demonstratives, if  $\models_T \phi$ , then  $\models_{LD} \phi$
- (ii)  $\models_{LD} \exists x$  Exist  $x$ , but  $\sim \models_T \exists x$  Exist  $x$

Of course ' $\Box \exists x$  Exist  $x$ ' is not valid even in LD. Nor are its counterparts, ' $\neg F \neg \exists x$  Exist  $x$ ', and ' $\neg P \neg \exists x$  Exist  $x$ '.

This suggests that we can transcend the context-oriented perspective of LD by generalizing over times and worlds so as to capture those possible circumstances  $\langle \mathcal{A}, t, w \rangle$  which do not correspond to any possible contexts  $\langle \mathcal{A}, c \rangle$ . We have the following result:

- (iii) If  $\phi$  contains no demonstratives,  
 $\models_T \phi$  iff  $\models_{LD} \Box (\neg F \neg \phi \wedge \neg P \neg \phi \wedge \phi)$ .

Although our definition of the neotraditional sense of validity was motivated by consideration of demonstrative-free formulas, we could apply it also to formulas containing essential occurrences of demonstratives. To do so would nullify the most interesting features of the logic of demonstratives. But it raises the question, can we express our new sense of validity in terms of the neotraditional sense? This can be done:

- (iv)  $\models_{LD} \phi$  iff  $\models_T AN\phi$

**Remark 10:** Rigid designators (in the sense of Kripke) are terms with a Stable Content. Since Kripke does not discuss demonstratives, his examples all have, in addition, a Stable Character (by Remark 8). Kripke claims that for proper names  $\alpha, \beta$  it may happen that  $\alpha = \beta$ , though not a priori, is nevertheless necessary. This, in spite of the fact that the names  $\alpha, \beta$  may be introduced by means of descriptions  $\alpha', \beta'$  for which  $\alpha' = \beta'$  is not necessary. An analogous situation holds in LD. Let  $\alpha', \beta'$  be definite descriptions (without free variables) such that  $\alpha' = \beta'$  is not a priori, and consider the (rigid) terms  $dthat[\alpha']$  and  $dthat[\beta']$  which are formed from them. We know that:

$$\models (dthat[\alpha'] = dthat[\beta'] \leftrightarrow \alpha' = \beta').$$

Thus, if  $\alpha' = \beta'$  is not a priori, neither is  $dthat[\alpha'] = dthat[\beta']$ . But, since:

$$\models (dthat[\alpha'] = dthat[\beta'] \rightarrow \Box (dthat[\alpha'] = dthat[\beta']))$$

it may happen that  $dthat[\alpha'] = dthat[\beta']$  is necessary. The converse situation can be illustrated in LD. Since  $(\alpha = dthat[\alpha])$  is valid (see Remark 3), it is surely capable of being known a priori. But if  $\alpha$  lacks a Stable Content (in some context  $c$ ),  $\Box(\alpha = dthat[\alpha])$  will be false.

**Remark 11:** Our 0-0-place  $i$ -functors are not proper names, in the sense of Kripke, since they do not have a Stable Content. But they can easily be converted by means of stabilizing influence of 'dthat'. Even  $dthat[\alpha]$  lacks a Stable Character. The process by which such expressions are converted into expressions with a Stable Character is 'dubbing'—a form of definition in which context may play an essential role. The means to deal with such context-indexed definitions is not available in our object language.

There would, of course, be no difficulty in supplementing our language with a syntactically distinctive set of 0-0-place  $i$ -functors whose semantics requires them to have both a Stable Character and a Stable Content in every context. Variables already behave this way, what is wanted is a class of constants that behave, in these respects, like variables.

The difficulty comes in expressing the definition. My thought is that when a name, like 'Bozo', is introduced by someone saying, in some context  $c^*$ , "Let's call the Governor, 'Bozo'", we have a context-indexed definition of the form:  $A =_{c^*} \alpha$ , where  $A$  is a new constant (here, 'Bozo') and  $\alpha$  is some term whose denotation depends on context (here, 'the Governor'). The intention of such a dubbing is, presumably, to induce the semantical clause: for all  $c$ ,  $\{A\}_{c^*}^{\mathcal{A}} = \{\alpha\}_{c^*}$ . Such a clause gives  $A$  a Stable Character. The context-indexing is required by the fact that the Content of  $\alpha$  (the 'definiens') may vary from context to context. Thus the same semantical clause is not induced by taking either  $A = \alpha$  or even  $A = dthat[\alpha]$  as an axiom.

I think it is likely that such definitions play a practically (and perhaps theoretically) indispensable role in the growth of language, allowing us to introduce a vast stock of names on the basis of a meager stock of demonstratives and some ingenuity in the staging of demonstrations.

Perhaps such introductions should not be called 'definitions' at all, since they essentially enrich the expressive power of the language. What a nameless man may express by 'I am hungry' may be inexpressible in remote contexts. But once he says "Let's call me 'Bozo'", his Content is accessible to us all.

**Remark 12:** The strongest form of logical equivalence between two formulas  $\phi$  and  $\phi'$  is sameness of Character,  $\{\phi\} = \{\phi'\}$ . This form of synonymy is expressible in terms of validity.

$$\{\phi\} = \{\phi'\} \text{ iff } \models \Box[\neg F\neg(\phi \leftrightarrow \phi') \wedge \neg P\neg(\phi \leftrightarrow \phi') \wedge (\phi \leftrightarrow \phi')]$$

[Using Remark 9 (iii) and dropping the condition, which was stated only to express the intended range of applicability of  $\models_T$ , we have:  $\{\phi\} = \{\phi'\}$  iff  $\models_T(\phi \leftrightarrow \phi')$ .] Since definitions of the usual kind (as opposed to dubbings) are intended to introduce a short expression as a mere abbreviation of a longer one, the Character of the defined sign should be the same as the Character of the definiens. Thus, within LD, definitional axioms must take the unusual form indicated above.

**Remark 13:** If  $\beta$  is a variable of the same sort as the term  $\alpha$  but is not free in  $\alpha$ , then  $\{\text{dthat}[\alpha]\} = \{\text{the } \beta \text{ AN}(\beta = \alpha)\}$ . Thus for every formula  $\phi$ , there can be constructed a formula  $\phi'$  such that  $\phi'$  contains no occurrence of 'dthat' and  $\{\phi\} = \{\phi'\}$ .

**Remark 14:**  $Y$  (yesterday) and  $G$  (one day ago) superficially resemble one another in view of the fact that  $\models (Y\phi \leftrightarrow G\phi)$ . But the former is a demonstrative whereas the latter is an iterative temporal operator. "One day ago it was the case that one day ago it was the case that John yawned" means that John yawned the day before yesterday. But "Yesterday it was the case that yesterday it was the case that John yawned" is only a stutter.

### Notes on Possible Refinements

1. The primitive predicates and functors of first-order predicate logic are all taken to be extensional. Alternatives are possible.
2. Many conditions might be added on  $\mathcal{P}$ ; many alternatives might be chosen for  $\mathcal{T}$ . If the elements of  $\mathcal{T}$  do not have a natural relation to play the role of  $<$ , such a relation must be added to the structure.
3. When  $K$  is a set of LD formulas,  $K \models \phi$  is easily defined in any of the usual ways.
4. Aspects of the contexts other than  $c_A$ ,  $c_P$ ,  $c_T$ , and  $c_W$  would be used if new demonstratives (e.g., pointings, You, etc.) were added to the language. (Note that the subscripts A, P, T, W are external parameters. They may be thought of as functions applying to contexts, with  $c_A$  being the value of A for the context  $c$ .)

5. Special continuity conditions through time might be added for the predicate 'Exist'.
6. If individuals lacking positions are admitted as agents of contexts, 3(iii) of page 543 should be weakened to:  $c_P \in \mathcal{P} \cup \{+\}$ . It would no longer be the case that:  $\models$  Located I, Here. If individuals also lacking temporal location (disembodied minds?) are admitted as agents of contexts, a similar weakening is required of 3(ii). In any case it would still be true that  $\models$  Exist I.

### XX. Adding 'Says'

[This section is not yet written. What follows is a rough outline of what is to come.]

The point of this section is to show, in a controlled experiment, that what Quine called *the relational sense* of certain intensional operators is unavoidable, and to explore the *logical*, as opposed to epistemological, features of language which lead to this result.

I have already mentioned, in connection with Dr. Lauben, that when  $x$  says 'I have been wounded' and  $y$  wishes to report in indirect discourse exactly what  $x$  said,  $y$  has a problem. It will not do for  $y$  to say 'x said that I have been wounded'. According to our earlier remarks, it should be correct for  $y$  to report  $x$ 's *content* using a character appropriate to the context of the report. For example, accusingly: 'You said that you had been wounded', or quantificationally: ' $(\exists z)(Fz \wedge x$  said that  $z$  had been wounded)' where  $x$  alone satisfied ' $Fz$ '. I will try to show that such constructions are the inevitable result of the attempt to make (third person) *indirect discourse* reports of the first person *direct discourse* sayings when those sayings involve indexicals.

The situation regarding the usual epistemic verbs—'believes', 'hopes', 'knows', 'desires', 'fears', etc.—is, I believe, essentially similar to that of 'says'. Each has, or might have, a *direct discourse* sense in which the character which stands for the cognitive significance of the thought is given (he thinks, 'My God! It is *my* pants that are on fire.') as well as an *indirect discourse* sense in which only the content need be given (he thinks that it is *his* pants that are on fire).<sup>68</sup> If this is correct, and if indexicals are featured in the language of thought (as suggested

<sup>68</sup>My notion of 'indirect discourse' forms of language is linked to Frege's notion of an 'ungerade' (often translated 'oblique') context. My terminology is intended to echo his.

earlier), then any *indirect* discourse reports of someone's thought (other than first person on the spot reports) must contain those features—*de re* constructions, referential occurrences, quantification in, relational senses—that have so puzzled me, and some others, since the appearance of "Quantifiers and Propositional Attitudes."<sup>69</sup>

What is special and different about the present approach is the attempt to use the distinction between direct and indirect discourse to match the distinction between character and content. Thus when you wonder, 'Is that me?', it is correct to report you as having wondered whether you are yourself. These transformations are traced to the indexical form of your inner direct discourse rather than to any particular referential intentions. The idea is that the full analysis of indirect discourse includes mention of the suppressed character of the direct discourse event which the indirect discourse reports, thus:

$\exists c, C [c \text{ is a context} \wedge C \text{ is a character} \wedge x \text{ is the agent of } c$   
 $\wedge x \text{ direct-discourse-verb } C \text{ at the time } t \text{ of } c \wedge \text{the content}$   
 $\text{of } C \text{ in } c \text{ is that...}]$

approximates a full analysis of

$x$  indirect-discourse-verb that ... at  $t$ .

Rather than try to include all these semantical ideas in an object language which includes the direct discourse forms of the verbs, the object language will include, *as is usual*, only the indirect discourse forms. The information about the character of the direct discourse event will provide the metalinguistic data against which the truth of object language sentences is tested.<sup>70</sup>

<sup>69</sup> Quine, in his "Reply to Kaplan" in *Words and Objections*, ed. D. Davidson et al. (Dordrecht: Reidel, 1969), raises the question—in the idiom of "Quantifiers and Propositional Attitudes" (*Journal of Philosophy* 53 (1956); reprinted in Martinich, op. cit.)—which of the names of a thing are to count as exportable? My point here is that the indexical names must be exportable, not because of some special justification for the transformation from a *de dicto* occurrence to a *de re* occurrence, but because indexicals are devices of direct reference and have no *de dicto* occurrences. I am reminded of the Zen ko-an: How do you get the goose out of the bottle? Answer: It's out!

<sup>70</sup> If this analysis is correct, the suppressed character should wreak its mischief in cases of suspension of belief (I believe, 'that man's pants are on fire' but at the moment neither assent to nor deny 'my pants are on fire') as does its counterpart in section XI of "Quantifying In." Burge, in "Kaplan, Quine, and Suspended Belief," *Philosophical Studies* 31 (1977): 197–203, proposes a solution to the problem of section XI which he believes is in the spirit of Quine's formulations. A similar

What is not yet clear to me is whether all directly referential occurrences of terms within the scope of indirect discourse epistemic verbs are to be justified *solely* on the basis of a like (though generally distinct) term in the direct discourse event or whether in some cases the English idioms which we symbolize with quantification in (for example, 'There is someone whom Holmes believes to have shot himself') involve some element of *knowing-who* or *believing-who*. To put the question another way: are all the cases that Quine describes, and others similar, which irresistibly suggest the symbolic idiom of quantification in, accounted for by the semantics of direct reference (including indexicals and possibly other expressions as well) as applied to the (putative) direct discourse events? "Quantifying In" suffers from the lack of an adequate semantics of direct reference, but its explicandum includes the epistemological idea of knowing-who, which goes beyond what can be analyzed simply in terms of direct reference. When Ingrid hears someone approaching through the fog and knows 'Someone is approaching' and even knows 'That person is approaching', is it justified to say that there is someone whom Ingrid knows to be approaching? Or must we have, in addition to the indexical 'that person', *recognition* on Ingrid's part of who it is that is approaching? My present thought is that the cases which irresistibly suggest the symbolic idiom of quantification in involve, in an ambiguous way, two elements: *direct reference* (on which we are close to getting clear, I hope) and *recognition*.<sup>71</sup> (The latter is my new term

proposal in the present context would seem starkly inappropriate. But there has been a shift in task from "Quantifying In" to the present attempt. In large part the shift is to a course outlined by Burge in the last two pages of the above-mentioned article and urged by him, in conversation, for several years. The point only began to sink in when I came on it myself from a different angle.

<sup>71</sup> There is another form of common speech which may be thought to suggest formalization by quantification in. I call this form the *pseudo de re*. A typical example is, "John says that the lying S.O.B. who took my car is honest." It is clear that John does not say, "The lying S.O.B. who took your car is honest." Does John say ' $\delta$  is honest' for some directly referential term  $\delta$  which the reporter believes to refer to the lying S.O.B. who took his car? Not necessarily. John may say something as simple as, "The man I sent to you yesterday is honest." The reporter has simply substituted his description for John's. What justifies this shocking falsification of John's speech? Nothing! But we do it, and often recognize—or don't care—when it is being done. The form lends itself to strikingly distorted reports. As Church has shown, in his *Introduction to Mathematical Logic* (Princeton: Princeton University Press, 1956), on page 25, when John says "Sir Walter Scott is the author of *Waverley*" use of the *pseudo de re* form (plus a quite plausible synonymy transformation) allows the report, "John says that there are twenty-nine counties in Utah"! I do not see that the existence of the *pseudo de re* form of report poses

for knowing-(or believing)-who.) The term is chosen to reflect the idea that the individual in question is identified with respect to some prior or independent information—*re-cognition*—not immediately connected with the current attribution.) Of the two elements the former is semantic; the latter, frankly epistemological. The English idiom ‘There is someone such that Ingrid indirect-discourse-propositional-attitude-verb that ...he ...’ always implies that a singular proposition is the object of Ingrid’s thought (and thus that some directly referential term  $\alpha$  occurred in her inner direct discourse) and may sometimes imply (or only suggest?) that Ingrid recognized, *who  $\alpha$  is*. I offer no analysis of the latter notion.<sup>72</sup>

In the first paragraph, I referred to a controlled experiment. By that I mean the following. Accepting the metaphor of “inner direct discourse events” and “indirect discourse reports” in connection with the usual epistemic verbs, I want to examine the logical relations between these two. But the study is complicated by at least three factors which obscure the issues I wish to bring to light. First, there is no real syntax to the language of thought. Thus, even in the case of the simplest thoughts the relation between the syntax of the sentential complement to the epistemic verb and the structure of the original thought is obscure. Second, in containing images, sounds, odors, etc., thought is richer than the language of the report. Might these perceptual elements play a role in determining logical relations? Third, thought ranges from the completely explicit (inner speech) to the entirely implicit (unconscious beliefs which explain actions) and through a variety of occurrent and dispositional forms. This makes it hard to pin down the whole direct discourse event. These three factors suggest taking as a paradigm of the relation between direct and indirect discourse—direct and indirect discourse!

Even when reporting the (outer) discourse of another, at least three obscure irrelevancies (for our purposes) remain. First, if Christopher speaks in a language different from that of the report, we have again the problem of translation (analogous to, though perhaps less severe than,

any issues of sufficient theoretical interest to make it worth pursuing.

<sup>72</sup>There is a considerable literature on this subject with important contributions by Hintikka, Castañeda and others. In connection with the proposal that ‘a knows who  $\alpha$  is’ can be symbolized ‘ $\exists x(a \text{ knows that } x = \alpha)$ ’, it should be noted that  $a$ ’s knowledge of the logical truth ‘ $\text{that}[\alpha] = \alpha$ ’ leads, simply by the semantics of direct reference, to ‘ $\exists x(a \text{ knows that } x = \alpha)$ ’. This shows only that a *recognition* sense of knowing a singular proposition is not definable, in the obvious way, in terms of a purely *direct reference* sense of knowing a singular proposition.

that of translating the language of thought). We control this by assuming the direct discourse to be in the language of the indirect discourse report. Second, as Carnap once pointed out to me, if Christopher’s discourse had the form ‘ $\phi \wedge \psi$ ’ even the strictest court would accept as true the testimony, ‘Christopher said that  $\psi \wedge \phi$ ’. What logical transformations on the original discourse would be allowed in the report? (If Christopher says ‘ $\exists x$   $x$  is round’, may we report him as saying that ‘ $\exists y$   $y$  is round?’) We control this by allowing no logical transformations (we are explicating *literal* indirect discourse). Third, if in saying ‘The circle can’t be squared’ Christopher thought that ‘can’t’ was synonymous with ‘should not’ rather than ‘cannot’, should he be reported as having said that the circle can’t be squared? We control this by assuming that our speakers make no linguistic errors.

What then remains of the logic? Is the move from direct discourse to literal indirect discourse not simply the result of disquotation (and decapitalization) plus the addition of ‘that’, as in:

Christopher says ‘the world is round’

∴ Christopher says that the world is round           ?

But how then are we to report Dr. Lauben’s saying, ‘I have been wounded’? Certainly not as, ‘Dr. Lauben says that I have been wounded’!

Even in this highly antiseptic environment, the logic of *says* should provide us with a full measure of that baffling and fascinating *de re* versus *de dicto*, notional versus relational, etc., behavior. And here, using the conceptual apparatus of the semantics of direct reference, we may hope to identify the source of these antics.

[I also hope to distinguish, in discussing reports of self-attribution, *x says that x is a fool*, from *x says-himself to be a fool*.]

## XXI. Russell on Egocentric Particulars and Their Dispensability

In chapter VII of *Inquiry Into Meaning and Truth*,<sup>73</sup> Russell gives a series of atrocious arguments for the conclusion that “[indexicals] are not needed in any part of the description of the world, whether physical or psychological.” This is a happy no-nonsense conclusion for an argument that begins by remarking “A physicist will not say ‘I saw a table’, but like Neurath or Julius Caesar, ‘Otto saw a table’.” [Why Julius Caesar would be provoked to say ‘Otto saw a table’, is unexplained.]

<sup>73</sup>Bertrand Russell (London: Allen & Unwin, 1940).

Let us examine Russell's conclusion without prejudice to his argument. [What follows is an outline.]

In brief, there are essentially two points. First: if we have both the indexicals and an unlimited supply of unused directly referential proper names, and we can do instantaneous dubbing, then in each context *c* for any sentence  $\phi$  containing indexicals we can produce a sentence  $\phi^*$  whose character is fixed and whose content is the same as that of  $\phi$  in *c*. In this sense, if you can describe it with indexicals you can describe it without.<sup>74</sup> There are problems: (i) things can change fast and dubbings take time, (ii) the indexicals retain a kind of epistemic priority.

The second point is: given any *prior* collection of proper names, there will be things, times, places, etc., without a name. How do I say something about these unnamed entities? (E.g., how do I tell you that your pants are on fire—now? It may be that nothing in sight, including us, and no nearby time has a name.)

There are two cases. It seems most likely that without indexicals some entities cannot even be uniquely *described*. In this case we are really in trouble (unless Russell believes in the identity of indescribables—objects lacking uniquely characterizing descriptions) because without indexicals we cannot freely introduce new names. If every entity *can* be uniquely described, there is still the problem of not presenting the right content under the right character required to motivate the right action (recall the discussion on pages 532–33). The proposition expressed by 'the pants belonging to *the x Fx* are on fire at *the t Gt*' is not the proposition I want to express, and certainly does not have the character I wish to convey.<sup>75</sup>

## XXII. On Proper Names

[Some thoughts on proper names from the perspective of the formal system are contained in Remark 11, page 551. What follows is the most hastily written section of this draft. I sketch a view that is mainly

<sup>74</sup>I assume here that proper names are not indexicals. I argue the point in section XXII.

<sup>75</sup>Some interesting arguments of a different sort for the indispensability of indexicals are given by Burge in "Belief De Re," *Journal of Philosophy* 74 (1977): 338–62, and by Bar-Hillel in his pioneering work, "Indexical Expressions," *Mind* (1954). In connection with the arguments of Burge and Bar-Hillel it would be interesting to check on some related empirical issues involving linguistic universals. Do all languages have a first person singular form? Do they all have all of the standard indexicals?

negative, without including much supporting argumentation (several of the omitted arguments seem both tedious and tendentious). My current inclination is to drop this whole section from the final draft.]

A *word* is an expression along with its meaning. When two expressions have the same meaning, as with "can't" and "cannot", we call the two words *synonyms*. When two meanings have the same expression, we call the two words *homonyms*. In the latter case we also say that the expression is *ambiguous*. (Probably we would say that the *word* is ambiguous, but accept my terminology for what follows.) In a disambiguated language, semantics can associate meanings with expressions. Even in a language containing ambiguities, semantics can associate a set of meanings with an expression. But given an utterance, semantics cannot tell us what expression was uttered or what language it was uttered in. This is a presemantic task. When I utter a particular vocable, for example, the one characteristic of the first person pronoun of English, you must decide what *word* I have spoken or indeed, if I have spoken any word at all (it may have been a cry of anguish). In associating a word with my utterance you take account of a variety of features of the context of utterance that help to *determine* what I have said but that need not be any *part* of what I have said. My egotism, my intonation, my demeanor, may all support the hypothesis that it was the first person pronoun of English. But these aspects of personality, fluency, and mood are no part of any semantic theory of the first person pronoun. The factors I have cited are not, of course, *critical* for the use of the first person pronoun. What are the criteria? What would definitively settle the question? I don't know. I think this is a very difficult question. But among the criteria there must be some that touch on the utterer's intention to use a word in conformity with the conventions of a particular linguistic community. For proper name words, in part because they are so easily introduced, this aspect of the presemantic determination is especially important.

According to the causal chain or chain of communication theory, there are two critical intentions associated with the use of the proper name word. One is the intention to use the word with the meaning given it by the person from whom you learned the word. The other is the contrary intention to create (and perhaps simultaneously use) a proper name word to refer to a given object irrespective of any prior meanings associated with the expression chosen as a vehicle. One who uses a proper name word with the first intention generally (but not always) believes that someone originated the word by using it with the



second intention, and—according to the causal chain theory—intends to refer to the given object.<sup>76</sup>

In “Bob and Carol and Ted and Alice,” appendix IX, I introduce the notion of a *dubbing* for what I took to be the standard form of introduction of a proper name word. That notion has been mistakenly taken to imply—what I deliberately sought to evoke—a formal public ceremony. What I actually had in mind was a use of a proper name word with the second intention: the intention to originate a word rather than conform to a prior usage. Thus a fleeting “Hi-ya, Beautiful” incorporates all the intentional elements required for me to say that a dubbing has taken place. I believe that my notion here is closely related to Donnellan’s notion of a *referential use* of a definite description. Donnellan’s distinction between referential and attributive uses of definite descriptions is easily and naturally extended to referential and attributive uses of proper names. When the intention to conform to a preestablished convention is absent we have the pure referential use. In this case, when a proper name is in question, I take it that an internal, subjective, dubbing has occurred. When a definite description is in question, again the speaker does not intend to give the expression its conventional meaning although he may intend to *make use* of the conventional meaning in conveying who it is that is being referred to or for some other purpose associated with

<sup>76</sup>There is disagreement as to how the given object must be given to one who introduces a proper name word with the second intention. Must he be acquainted with the object, directly acquainted, *en rapport*, perceiving it, causally connected, or what? My liberality with respect to the introduction of directly referring terms by means of ‘dthat’ extends to proper names, and I would allow an arbitrary definite description to give us the object we name. “Let’s call the first child to be born in the twenty-first century ‘Newman 1.’” But I am aware that this is a very controversial position. Perhaps some of the sting can be removed by adopting an idea of Gilbert Harman. Normally one would not introduce a proper name or a dthat-term to correspond to each definite description one uses. But we have the means to do so if we wish. Should we do so, we are enabled to apprehend singular propositions concerning remote individuals (those formerly known only by description). Recognizing this, we refrain. What purpose—other than to confound the skeptics—is served by direct reference to whosoever may be the next president of Brazil? The introduction of a new proper name by means of a dubbing in terms of description and the active contemplation of characters involving dthat-terms—two mechanisms for providing direct reference to the denotation of an arbitrary definite description—constitute a form of cognitive restructuring; they broaden our range of thought. To take such a step is an action normally not performed at all, and rarely, if ever, done capriciously. The fact that we have the means—without special experience, knowledge, or whatever—to refer directly to the myriad individuals we can describe, does not imply that we will do so. And if we should have reason to do so, why not?

the act of utterance (as in “Hi-ya, Beautiful”). What is important here is that the speaker intends to be creating a meaning for the expression in question rather than following conventions. Dubbings, whether aimed at introducing a relatively permanent sense for the expression or only aimed at attaching a nonce-sense to the expression, are unconventional uses of language. Dubbings create words.

In many, perhaps most, uses of definite descriptions there is a mixture of the intention to follow convention with the intention to refer to a preconceived individual. The same mixture of ‘attributive’ and ‘referential’ intentions can occur with a proper name. If I introduce a name into your vocabulary by means of false introduction (“This is Jaakko Hintikka”, but it isn’t), you are left with an undiscriminated tangle of attributive (to refer to Jaakko Hintikka) and referential (to refer to the person to whom you were introduced) intentions associated with your subsequent uses of the expression ‘Jaakko Hintikka’. There are several ways in which one might attempt to account for these mixed intentions in a general theory of language. First, we might distinguish two notions: speaker’s-reference and semantic-reference. The presence of an attributive intention justifies giving the expressions a conventional meaning and thus allows us to claim that preexisting words were used. Whereas the presence of a referential intention (not just a *belief* that the semantic referent is the given object, but an independent intention to refer to the given object independent of any particular interpretation of the expressions he used as words and independent of whether the utterance has an interpretation as words. A second way of accounting for mixed intentions of this kind is to assume that one of the two intentions must be dominant. If the referential intention dominates, we regard the utterance, on the model of “Hi-ya, Beautiful,” as an apt (or inept, as the case may be) introduction of a proper name word (or phrase). Thus, as essentially involving a dubbing. On this way of accounting for mixed intentions, a referential use of an expression would endow the expression with a semantic referent identical with the speaker’s referent.<sup>77</sup>

<sup>77</sup>This is not an unnatural way to account for the use of the proper name word in the false introduction case, but it does seem a bit strange in the case of a definite description. In that case it involves hypothesizing that the speaker intended the description expression to have a meaning which made the given object its semantic referent, and only *believed* that the conventional meaning would do this, a belief that he is prepared to give up rather than acknowledge that the semantic referent of his words was not the given object. Something like this seems to happen when descriptions grow capitals, as in ‘The Holy Roman Empire’, and in other cases as

My aim in the foregoing is to emphasize how delicate and subtle our analysis of the context of utterance must be for the presemantic purpose of determining what words, if any, were spoken. I do this to make plausible my view that—assuming the causal chain theory of reference—proper names are not indexicals. The contextual feature which consists of the causal history of a particular proper name expression in the agent's idiolect seems more naturally to be regarded as determining what word was used than as fixing the content of a single context-sensitive word. Although it is true that two utterances of 'Aristotle' in different contexts may have different contents, I am inclined to attribute this difference to the fact that distinct homonymous words were uttered rather than a context sensitivity in the character of a single word 'Aristotle'. Unlike indexicals like 'I', proper names really are ambiguous. The causal theory of reference tells us, in terms of contextual features (including the speaker's intentions) which word is being used in a given utterance. Each such word is directly referential (thus it has a fixed content), and it also has a fixed character. Therefore, in the case of proper name words, all three kinds of meaning—referent, content, and character—collapse. In this, proper name words are unique. They have the direct reference of indexicals, but they are not context-sensitive. Proper name words are like indexicals that you can carry away from their original context without affecting their content. Because of the collapse of character, content, and referent, it is not unnatural to say of proper names that they have no meaning other than their referent.

Some may claim that they simply use 'indexical' in a wider sense than I (perhaps to mean something like 'contextual'). But we must be wary of an overbroad usage. Is every ambiguous expression an indexical because we look to utterer's intentions to disambiguate? Indeed, is every expression an indexical because it might have been a groan?

If the character and content of proper name words is as I have described it (according to the causal theory), then the informativeness of  $\lceil \alpha = \beta \rceil$ , with  $\alpha$  and  $\beta$  proper names, is not accounted for in terms of differences in either content or character. The problem is that proper names do not seem to fit into the whole semantical and epistemological scheme as I have developed it. I claimed that a competent speaker knows the character of words. This suggests (even if it does not imply) that if two proper names have the same character, the competent speaker

knows that. But he doesn't. What is perhaps even more astounding is that I may introduce a new proper name word and send it on its journey. When it returns to me—perhaps slightly distorted phonologically by its trip through other dialects—I can competently take it into my vocabulary without recognizing it as the very same word! Shocking!

In earlier sections of this paper I have tried to show that many of the metaphysical and epistemological anomalies involving proper names had counterparts involving indexicals, and further that in the case of indexicals these wonders are easily explained by an obvious theory. Insofar as I am correct in regarding the anomalies as counterparts, the theory of indexicals may help to break down unwarranted resistance to the causal chain theory. It may also suggest the form of a general semantical and epistemological scheme comprehending both indexicals and proper names. This is not the place to attempt the latter task; my purpose here is simply to show that it is not trivial.<sup>78</sup> Those who suggest that proper names are merely one species of indexical depreciate the power and the mystery of the causal chain theory.

<sup>78</sup>The issues to be resolved by "a general semantical and epistemological scheme comprehending ... proper names" are such as these. Is the work of the causal chain theory presemantic, as I have claimed? Do proper names have a kind of meaning other than reference? Does the causal chain theory itself constitute a kind of meaning for proper names that is analogous to character for indexicals (but which, perhaps, gives all proper names the same meaning in this sense)? Are proper names words of any particular language? Is there synonymy between proper names that are expressed differently (as there is between 'can't' and 'cannot')? How should we describe the linguistic competence of one who does not know that Hesperus is Phosphorus? Is he guilty of linguistic error? Should we say he does not know what words he speaks? Does he know that 'Hesperus' and 'Phosphorus' are different words? Are they? Is it really possible, as I claim, to account for the semantics of indexicals without making use of the full conceptual resources required to account for the semantics of proper names? I raise these issues—and there are others—within the framework of a hypothetical acceptance of the causal chain theory. There are other issues, of a quite different kind, involved in trying to fill out some details of the causal chain theory itself. For example, if one who has received some particular proper name expression, say, "James", hundreds of times, uses that expression attributively as a proper name, and has in mind no particular source, how do we decide which branch to follow back? The first set of issues seems to me to be largely independent of the details of the relevant causal chains.

well, for example Russell's 'denoting phrases' which do not denote. But it still seems strange.