RECURSIVE SEMANTICS FOR KNOWLEDGE AND BELIEF

1. This paper is an informal exposition of a model-theoretic semantics for knowledge and belief set out in full detail else where.[1] Considerations of space and simplicity prevent any recapitulation of tracts of formal definitions. My aim is simply to inform the reader of the alleged existence of one "new direction" in semantics, and to direct him to the original source for its detailed development. I shall explain certain self-imposed limitations on the scope and adequacy conditions of this treatment. Then, in subsequent discussion of the semantical problems which I claim to have treated, I shall try to impart the flavour of the technical methods involved.

For reasons which will be set out elsewhere in a longer exegetical and critical study of what may be called the Hintikka paradigm in the logic of knowledge and belief [2], the adequate semantical treatment of propositional attitudes must consist of a truth definition along the lines of that of Tarski for classical first order languages. Not sharing Davidson's opposition to model-theoretic versions of truth definitions[3] my basic premiss is that a correct semantics for propositional attitudes should show how the truth or falsity of attitude reports is a matter arising from attitude schemes (or their formal representations) *vis à vis* suitably chosen canonical sentences representing the logical forms of those reports on particular readings. Just as a formula of elementary number theory is true or false *vis à vis* the particular relational structure in which it is interpreted, so in the system I propose a formula involving operators corresponding to attitude verbs will be true or false *vis à vis* the *attitudinal model* in which it is interpreted.

So firstly, different readings (*de re, de dicto*) of, say, belief reports, must be captured by different canonical sentences from our logical grammar; and secondly, we must find some way of modelling people's belief schemes about the world and about others' belief schemes about the world. . . and so on.

2. In facing these two problems we simplify matters by considering only attitude reports whose subordinate clauses are of a simple syntactic and semantic structure. We cannon embark simultaneously on a treatment of mood, aspect, tense, adverbial modification etc. My aim is to treat only those subordinate clauses which are complicated enough to display the puzzles

which have so far been identified as important for a semantics for propositional attitudes, but which are simple enough to fall within a self-contained and general account which concerns itself fully with just those problems.

We shall imagine predicates of the language as holding of publicly observable objects, and we shall imagine speakers of the language as coming to believe about them such things as are expressible in the language. An attitudinal model will then contain a representation of the actual state of affairs, and of the belief schemes of the speakers, at any given time. Despite these self-imposed limitations all the problems which have been raised in the literature arise within this framework, so simplification of grammar and limitation of vocabulary do not yield commensurate semantical simplification so far as attitude verbs are concerned.

The limitation of vocabulary bears directly upon how deep a logical analysis must be given of any sentence reporting knowledge or belief. Martin[4] maintains that we should analyze them maximally, "bringing out all the structural or grammatical components of the sentence expressing what is believed." This can be done when we are dealing with language systems (such as those of Carnap)[5] in which the primitives are always given. But, Martin asks,

> What are the primitives of a natural language? Well, strictly it would seem, there are none. But in a given context or for given purposes, some terms may presumably be picked out and regarded as primitives. One's analysis would then be relative to those selected terms. A maximum analysis in one context need not be a maximum analysis in another.

Compare Quine:[6] "Whether a general term in a sentence of ordinary language survives in a canonical paraphrase of the sentence, or disappears in favor of a more minute analysis, depends only on one's momentary purposes in paraphrasing." With Quine, however, this observation is linked to his "maxim of shallow analysis", according to which no more sentential structure need be revealed than is needed in order to appraise the logical relationships which hold one's momentary interests. This observation on its own does not rule out the possibility of being able to stipulate what expressions of the language are to be taken as primitive in a given context. There may not be

> an absolute vocabulary of simple general terms, as canonical all-purpose elements of paraphrase;

but there is no reason to suppose that we cannot, relative to a fixed context

and within a simple fragment of natural language, draw on a lexicon of primitives in giving full analyses of the structure of sentences *in that context*. From every vantage point the web of belief, as it were, must have its finest strands.

The final limitation on the scope of our treatment is, arguably, not a limitation at all. Our semantics confers upon the canonical sentence aKp the sense

> a knows, or is logically committed by the truth of what he knows, to the truth of the claim that p;

and upon aBp the sense

> a (consistently) believes, or is logically committed by the truth of what he (consistently) believes, to the truth of the claim that p.

(Knowledge is consistent; beliefs need not be, which is the reason for the parenthetical qualification here).[7] Of course we then encounter the usual "faithful descriptivist" objection that we are disregarding actual believers' logical myopia and proneness to inconsistency. This objection is correct but misdirected. First, it is "know" and "believe" in their ordinary senses which appear in the readings given above. K and B thereby inherit all the perplexing semantical properties of "know" and "believe". Secondly, the closer the ordinary verbs come to having the senses of the corresponding operators, the better our body of knowledge and beliefs will be from the logical point view. This much is tacitly admitted when the derivability of a contradiction from a man's presently held, explicitly stated or occurent beliefs is taken as an indication that his system of beliefs is unsound. So I have framed the semantics in such a way as to yield a logic of committal, and have left as a problem for further research that of modifying criteria for committal so as to have belief "grading off " over deductive distance. Suffice it to say that the modified criteria must be expressed in such a way as to be invariant from one system of deduction to another.

3. Our account, then, proceeds by simply adding the two operators K and B to the vocabulary of first order logic, incorporating them straight-forwardly into the definition of well formed formula. This modest addition to syntax induces problems of semantical interpretation including all those which have been raised in the literature. It seems premature, however—and, I hope to have shown, incorrect—to say that intensional contexts will thwart any attempt to provide a truth definition for a language whose sentences contain such contexts.[8] For we may introduce a new dimension of recursiveness into the truth definition. In the classical case, models are defined straight-

forwardly, formulae inductively, and truth-in a-model inductively over the well-founded structure of formulae. In our account of propositional attitudes *we now define models inductively as well (building them up in a stage by stage fashion)* and then define truth-in-a-model inductively over the *joint* well-founded structure of model and formula. In this way we can overcome the problem of accounting for *iterations of the attitude verbs.*

4. It would be helpful, in setting forth our semantics, if there were as perspicuous a relation as possible between English sentences reporting knowledge and belief on the one hand, and on the other, *the formulae proposed as capturing their logical forms. Unfortunately, as is well known, there are* no clear and reliable surface grammatical indicators of that semantical property which is of most interest in connection with knowledge and belief reports. This property is the *de re* or *de dicto* status of the reported knowledge or belief—and, derivatively, of the reports themselves. Henceforth I shall confine myself to the case of belief reports.

The de re-de dicto distinction will be familiar to the reader and I shall not need to impart a grasp of it by means of examples. (The distinction has gone under other headings in the literature: "Transparent-opaque", "relational-notional" and "objectual-propositional".) The status of a reported belief as *de re* or *de dicto* is best determined by "probabilities conferred by context", to use a phrase of Dummett. It depends on pragmatic considerations of context of utterance, and of speaker's, hearers' and discussees' background and sphere of acquaintance. Writers on these topics have used various devices in "logicians' English" to mark the distinction, such as the consistent use of the passive voice for *de re* and the active voice for *de dicto* readings:

j is believed by a to G	a believes that j is G
ıx(j = x,aBGx)	aBGj
The F is believed by a to G	a believes that the F is G
ıx(Fx, aBGx)	aBıx(Fx,Gx)
Someone is believed by a to G	a believes that someone is G
∃x aBGx	aB∃xGx
Everyone is believed by a to G	a believes that everyone is G
∀x aBGx	aB∀xGx

In all these cases the left-right ordering of the verb "believe" and the name or quantifying phrase in the English sentence is reflected directly in the cor-

responding formulae by the relative scopes of the operator B and the appropriate variable-binding operator. Unfortunately, however, with English sentences as they are actually used the relationship between *surface* and logical form is not in general this direct.

In particular, in the case of the so-called singular belief reports

(1) a believes that j is G

(2) a believes that the F is G

we have not yet uncovered all the readings. Consider the example of Alfred who believes that his wife Jane is gardening. Of Jane he believes (*de re*) that *she is gardening*; moreover he may well express this belief by saying "Jane is gardening". *Now our logical syntax provides for such a possibility by allow-*ing the following hybrid logical form for "Alfred believes that Jane is gardening":

(3) ⌐x(j = x,aB(j = x&Gx)).

This is the logical form of the belief report taken on the reading just explained. The reporter refers to the person about whom the belief is held and uses a certain name (here "Jane") in doing so; furthermore, he tacitly attributes to the believer the use of the same name to refer to that person should he ever express his belief about her. This mixed reading of the belief report may be expressed in a more roundabout way as

Alfred believes of Jane, as Jane, that she is gardening.

English speakers usually do not choose sentence forms which make *de re* or *de dicto* construals of their belief reports clear. Indeed, reports such as (1) and (2) are commonly construed as combining the two readings into the sort of mixed reading just explained. For this reason, everyday singular belief reports are governed by semantical intuitions which are best explained by assigning them the "hybrid" logical forms suggested. Linguists often talk of "forcing" an opaque or transparent reading on such sentences. If we take, for example,

(2) a believes that the F is G

we can force the transparent reading which has logical form

(4) ⌐x(Fx,aBGx)

instead of the hybrid form

(5) ⌐x(Fx,aB(⌐y(Fy,y = x&Gx)).

We can also force the opaque reading with logical form

$$aBｱx(Fx,Gx)$$

instead of the hybrid form. I contend that logical forms of English sentences should be regarded as logical forms of those sentences *on given readings.* If the logical representations given above are then taken as the deep structures underlying the singular belief report (2) on the respective readings, we shall be in the position of having the deep structures contain all the structural information necessary in order to determine the reading in question. A generative semanticist *might then look for admissible sequences of transformations in an English grammar which would map the three deep structures to* (2), thereby making the latter at least triply ambiguous.

If this view of the matter is correct, it would explain the difficulties which have stood in the way of an account of the logical behaviour of so-called singular terms (i.e. names or descriptive phrases) in the subordinate clauses of belief reports. The single occurrence of "the F" in (2) has led in the past to an overhasty representation of the logical form of (2) as aBG(the F). Difficulties then arise as to how "the F" can fulfil its referring function and yet not obey the law of substitutivity of identicals at that occurrence in the subordinate clause.

On the present view, however, "that occurrence" is a chimera. Sometimes, as in the mixed reading, there are really two occurrences at the level of deep structure or logical form; or, if there is only one occurrence at this level, then it may be not within the scope of B, but rather outside it, as is the case with the *de re* reading. Only on the *de dicto* reading is there just one occurrence within the scope of B. Surface form, as ever, is here a poor guide to logical form.

5. In providing logical forms like (4) and (5) for apparently singular belief reports we are already "quantifying into" belief contexts. For example, by representing the logical form of (2) on the *de re* reading by

$$(4)\ ｱx(Fx,aBGx)$$

we are using ｱ to quantify into the belief context. ｱ binds the variable x in the open formula Gx which is immediately subordinate to, or within the scope of B.

In general, quantifying in turns on the belief's being *de re.* Quantification objectually construed is explained in terms of the notion of satisfaction; and for quantification into belief contexts we extend this notion to open sentences of the form xBGy:

the sequence ⟨a,b⟩ satisfies xBGy
if and only if
a believes b *(de re)* to G.

It is not my purpose here to provide a noncircular philosophical analysis of *de re* belief proceeding linearly from clear, primitive notions to more difficult ones. I am doubtful whether such an analysis of *de re* belief is available. I shall only point out here the following features of *de re* belief:

(i) In order that his belief about b be *de re*, a must have been acquainted (in the common, non-Russellian sense) with b;

(ii) it is not necessary that a be able successfully to re-identify b on any subsequent occasion (consider, for example, Ralph's *de re* beliefs about Ortcutt after his different encounters with him in Quine's well-known story);

(iii) nor need a's belief about b be formed on any occasion of acquaintance with b; it is necessary only that the belief concern b as remembered from one or more occasions of acquaintance.

Now this relation of acquaintance which is necessary for the attribution of *de re* belief is philosophically problematic. Especially difficult is the task of extending the relation, understood as holding between persons and physical objects, to one holding also between persons and abstract objects such as numbers. I do not intend to investigate whether it makes sense to attribute *de re* belief about numbers, nor—if this does make sense—whether the relation of acquaintance may be extended to support such attributions. I shall explicitly limit myself to the case of *de re* belief about publicly observable objects.

6. Whatever the details of delineation of *de re* belief in this case—whatever causal interactions or chains between a believer and an object make his belief *de re*—I am concerned only to provide a schematic representation of the semantical outcome. The result of such interactions will, at any given time, be an articulable, mental representation on the part of the believer of the domain of objects with which he has been acquainted. The elements of such a representation will be impressions (in the common, non-Humean sense) of those objects. For the purposes of framing a formal semantics of belief I shall call these impressions *counterparts* (in a non-David Lewis sense!) *in* the person's belief scheme, *of* objects in the real world.

These objects are durable, observable and perhaps movable. A person may be acquainted with some of them on one or more occasions. As a result he may make some incorrect identifications and fail to make (even disbelieve) some correct identifications. With respect to such a person the counterpart relation on its remaining two arguments will be many-many. For example, a

might have met b twice and d once, and mistakenly identified b the second time with d. Within the belief scheme assigned to a the following situation would then obtain:
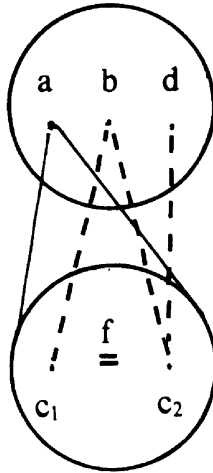

(In this diagram, = is the identity symbol; f is the truth value false. The top circle represents the real world, and the lower circle, assigned to a, represents a's belief scheme. As indicated by the dotted lines, c_1 and c_2 are counterparts, in a's belief scheme, of b; and c_2 is also a counterpart of d.)

This model would satisfy the following formulae:
$\exists x \exists y \exists z(\sim y = z$ & $xB(y = z))$
$\exists x \exists y \exists z(y = z$ & $B(\sim y = z))$.
A simple case of singular *de re* belief might as follows:
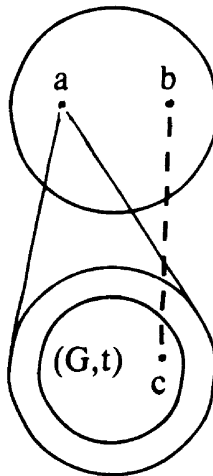
Here a counterpart c of b falls in the t-extension of the atomic predicate G. a believes b (*de re*) to G.

Formally, belief schemes are three-valued models employing the values t, f and u in order to represent belief, disbelief and (neither belief nor disbelief) respectively. The formal details ensure, via a method akin to that of supervaluations or model-completions, that a person may

(i) believe a disjunction without believing either of the disjuncts;

(ii) believe an existential quantification without believing (*de re* or *de dicto*) any instance;

(iii) believe of each F that it is both an F and a G without believing that all F's are G's; and

(iv) both believe of an object that it F's and believe of the same object that it does not F, without believing of that object that it both F's and does not F.

Since belief-schemes are formally three-valued, it is appropriate to talk of t-, u- and f-satisfaction therein. The crucial new clause in the definition of satisfaction is

$\langle a, \vec{b} \rangle$ t-satisfies $xBG\vec{y}$ in M

iff

there is a pointwise counterpart sequence $\vec{c}$ for $\vec{b}$ in the domain of the belief scheme M_a assigned to a in M such that $\vec{c}$ t-satisfies $G\vec{y}$ in M_a; and

$\langle a, \vec{b} \rangle$ f-satisfies $xBG\vec{y}$ in M

iff

either some member of $\vec{b}$ has no counterpart in the domain of M_a *or* some counterpart sequence of $\vec{b}$ u-satisfies or f-satisfies $G\vec{y}$ in M_a.

Thus attribution of *de rebus* belief is false either if the believer is not acquainted with all the objects concerned; or, when he is, if he either disbelieves or neither believes nor disbelieves the predication to hold.

Nowhere in our account is there any mention of the "possible worlds" of contemporary intensional semantics. We do employ a notion of *extension* or *completion* of a partly undecided system of beliefs which is a logical construct designed to help settle the question of what follows logically from a certain set of beliefs. Thus a man is committed to anything which is not f-satisfied by any extension of his belief scheme. There is no need, however, to construe an

extension in this sense as a possible world enjoying some problematic mode of existence alongside the actual world.

Nor are counterparts analogous to unactualized or unactualizable possibilia. They are simply formal correlates of mental constructs. They are a modest addition to ontology given what a theory of belief is trying to treat. Fleeting glimpses, detailed examinations, cursory inspections and long-standing acquaintance all give rise to what we ordinarily call faint, accurate, mistaken or lasting impressions upon us. In giving a formal semantics of belief statements—which describe, as Russell put it, "the most 'mental' thing we do, the thing most removed from what is done by mere matter"—I introduce counterparts as formal correlates of these mental impressions. They are not to be understood as momentary sense data. Friends and members of my family have over the years given rise to a stable system of counterparts in my belief scheme; as have the familiar objects on my study table, after many a long period of contemplation. As I learn more about those people and things, new predicates accrue to their counterparts in my belief scheme.

The more accurate and comprehensive a person's beliefs, the closer his belief scheme and the real world come to isomorphism. Counterparts sustain the de re predications which a person believes. Each counterpart may sustain anything from a rather meagre to a very rich cluster of predicates, depending on what sort of object (or objects) it is a counterpart of, and depending also on the extent to which the believer is acquainted with the object (or objects) responsible for his having such an impression as the counterpart represents.

Counterparts help to account for the structure of de re beliefs about the objects of our acquaintance. If a believer is consistent, then his de re beliefs can be reconciled with one another by an appropriate relational structure over a domain of counterparts suitably related to objects of his acquaintance. There may, of course, be more than one way of doing this for a "non-categorical" set of beliefs; but that is an epistemological, not a logical problem.

7. The relationship between knowledge schemes and belief schemes and between knowledge schemes and the real world may be secured in the formal semantics by requiring certain homomorphic embeddings to obtain. The consistency of problematic sentences such as

Someone is such that a believes that he does not exist

may be secured by intorducing quantifiers with a substitutional interpretation. "Knowing who" locutions, I would contend, involve second order quantifications and are therefore outside the scope of the present, first order treatment. First order logical omniscience ("What a knows to be F's he knows to be all the F's there are") may be treated by a suitable addition to the formal

vocabulary; as can Castañeda's distinction between self-conscious and non-self-conscious belief about oneself.[9] Moreover, we allow quantification over argument places for believers and knowers, which is a feature lacking from Hintikka's system in *Knowledge and Belief.*

Most important, however, is that our theory gives an account of the intelligible iterability of the attitude verbs. Attitudinal models are defined inductively by levels. If a sentence has n iterations of attitude verbs one within the scope of another then for its interpretation we need a model of level $(n+1)$. In a second level model, for example, believers are assigned first level models as belief schemes, so that the second level model can represent a's beliefs about b's beliefs about c. The definitions of model and of satisfaction-in-a-model are recursively interdependent; but the form of the satisfaction clause given above for sentences of the form xBGy is invariant under change of level. Thus the same analysis of B is provided at each of its occurrences in iterative contexts. This novel feature of our account removes an old obstacle to the semantics for propositional attitudes.

Neil Tennant

University of Edinburgh

NOTES

1. N. Tennant, "Recursive Semantics for Knowledge and Belief" (Ph.D. diss., University of Cambridge, 1974).

2. Jaakko Hintikka, *Knowledge and Belief* (Ithaca: Cornell University Press, 1967); N. Tennant, *The Hintikka Paradigm* (forthcoming).

3. D. Davidson, "In Defense of Convention T" in *Truth, Syntax and Modality,* ed. H. Leblanc (Amsterdam: N. Holland, 1973).

4. R.M. Martin, "On Knowing, Believing, Thinking," *Journal of Philosophy* 59 (1962).

5. R. Carnap, *Meaning and Necessity* (Chicago: University of Chicago Press, 1956).

6. W.V. Quine, *Word and Object* (Cambridge, Mass.: MIT Press, 1960), p. 204.

7. The precise explication of the adverbal "(consistently)" is rather involved. A first attempt to make our reading of aBp more exact might be (1) a's beliefs are consistent and p follows from them. But if $\vdash$p, then $\vdash$aBp and p follows from a's beliefs (whatever the latter); therefore a's beliefs must be consistent—which means that "belief" does not have its ordinary sense in (1).

A second attempt might be (2) *either* $\vdash$p *or* p follows from a's beliefs. But $\vdash$-aB(p&-p), and the reading (2) does not justify this. In order, therefore, to be sure that we have captured the force of B on the "committal" interpretation, and preserved the ordinary sense of "belief" in doing so, we should read aBp as *either* a's beliefs are inconsistent and $\vdash$p *or* a's beliefs are consistent and p follows from them. (I am indebted here to Dr. T. J. Smiley and Dr. K. Fine.)

8. Thus Quine: ". . . intensional contexts. . . block the recursion of a Tarskian truth definition," "Reply to Davidson," in *Words and Objections*, ed. D. Davidson and Jaakko Hintikka (Dordrecht-Holland: D. Reidel, 1969).

9. H.-N. Castañeda, " 'He': A Study in the Logic of Self-consciousness," *Ratio* 7 (1966).