Multiple quantifiers and restricted range in epistemic logic.

Gregory Frank Mellema
University of Massachusetts Amherst

Follow this and additional works at: http://scholarworks.umass.edu/dissertations_1

Recommended Citation
http://scholarworks.umass.edu/dissertations_1/2169

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations 1896 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
MUILTE O UANTIFIERS AND RESTRICTED RANGE
IN EPISTEMIC LOGIC

A Dissertation Presented
By
Gregory Frank Mellema

Submitted to the Graduate School of the
University of Massachusetts in partial
fulfillment of the requirements for the degree of
PHILOSOPHY
July 1974

Philosophy
MULTIPLE QUANTIFIERS AND RESTRICTED RANGE
IN EPISTEMIC LOGIC

A Dissertation Presented
By
Gregory Frank Mellema

Approved as to style and content by:

[Signatures]

Robert Sleigh, Chairman of Committee

Vere Chappell, Department Head

Edmund Gettier, Member

Terence Parsons, Member

Barbara Partee, Member

July, 1974
ABSTRACT

Most noteworthy among the developments in Epistemic Logic during the last five years are Hintikka's proposed semantics for perception (modelled after his 1962 semantics for knowledge) and his revised (1970) semantics for knowledge. Crucial to both of these systems is the employment of two sets of quantifiers. Part One of this dissertation explores the role of multiple quantifiers in Hintikka-type semantics. Chapter One is an investigation of Hintikka's semantics for perception; Chapter Two is a critique of a recent (1973) variation of Hintikka's semantics for perception (which makes use of multiple quantifiers) proposed by R. Thomason; And in Chapter Three Hintikka's 1970 semantics for knowledge is explored and criticized.

Part Two of this dissertation deals with the much-criticized Restricted Range feature which characterizes all of Hintikka's systems. Chapter Four is an examination of problems Restricted Range has been thought to create and a survey of some attempts to create Hintikka-type systems which lack the feature. Finally, in Chapter Five a new variation of Hintikka's semantics for knowledge is proposed which both lacks Restricted Range and avoids a number of troublesome theorems which show up in the other systems.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>v</td>
</tr>
<tr>
<td>Chapter One</td>
<td>1</td>
</tr>
<tr>
<td>Chapter Two</td>
<td>38</td>
</tr>
<tr>
<td>Chapter Three</td>
<td>67</td>
</tr>
<tr>
<td>Chapter Four</td>
<td>102</td>
</tr>
<tr>
<td>Chapter Five</td>
<td>133</td>
</tr>
<tr>
<td>Bibliography</td>
<td>166</td>
</tr>
</tbody>
</table>
PREFACE

In 1962, with the publication of Knowledge and Belief, Jaakko Hintikka introduced formal model set theoretic semantics for both knowledge and belief. The technique he employed consists of introducing a series of rules determining membership of well-formed formulas in model sets in such a way that model sets can be thought of as (possibly partial) descriptions of possible worlds. By introducing the notion of 'compatibility' among model sets, he was able to treat the symbols 'Ka' (read "a knows that") and 'Ba' (read "a believes that") as modal operators. Thus, the formula 'Kap' belongs to a model set (is true in a world) just in case 'p' belongs to every compatible model set (is true in every compatible world) relative to agent a. Due largely to Hintikka's proposal to treat epistemic terms as modal operators and at the same to allow quantifying in, Hintikka's systems have been the objects of widespread controversy and discussion ever since they first appeared.

Much of the early criticism of these systems which appeared was either mistaken or based upon misunderstanding of Hintikka's proposals, but genuine defects of Hintikka's semantics for knowledge (called "KB") were uncovered by Castaneda and Sleigh during the years 1964-1967. A good deal of criticism, both genuine and mistaken, was leveled at the so-called "Restricted Range" feature of KB. In his 1967 paper, 'Individuals, Possible Worlds, and Epistemic Logic', Hintikka proposed a significant revision of KB to avoid certain objections raised by Castaneda and Sleigh. The resulting system, IPE, however did not lack the Restricted Range feature; in the same paper
Hintikka defended Restricted Range and argued that it's unavoidable in his systems.

At about the same time Hintikka began to formulate a model set theoretic semantics for perception significantly more complex than his previous systems. Like the logics of knowledge and belief, the logic of perception is believed by Hintikka to be a "branch of modal logic", but he feels that to capture the complexities of the logic of perception it is necessary to employ two sets of quantifiers in the language of the system. In this system two independent methods are employed according to which individuals are traced across possible worlds. Hintikka's semantics for perception was first presented in his 1969 paper, 'On the Logic of Perception'.

The next year, 1970, Hintikka made the proposal (in the face of continued criticism by Castaneda) that two sets of quantifiers likewise be employed in his semantics for knowledge. Convinced that the logic of knowledge is more complex that his previous systems indicated, Hintikka dropped IPE and outlined a new system KBC in his paper 'On Attributions of Self-Knowledge'. As in his semantics for perception, there are two independent methods employed in KBC to trace individuals across possible worlds.

The present study divides into two parts. The first three chapters deal with multiple quantifiers in Hintikka-type systems. Chapter One is an investigation of Hintikka's semantics for perception; Chapter Two is a critique of a recent (1973) variation of Hintikka's semantics for perception (which makes use of multiple quantifiers) proposed by R. Thomason; And in Chapter Three Hintikka's 1970 semantics for knowledge, KBC, is explored and criticized. The last
two chapters deal with Restricted Range. Chapter Four is an examination of problems Restricted Range has been thought to create and a survey of some attempts to create Hintikka-type systems which lack the feature. Finally, in Chapter Five a new variation of KB is proposed which both lacks Restricted Range and avoids a number of troublesome theorems which show up in the other systems.
CHAPTER I

In his papers 'On the Logic of Perception' and 'Information, Causality, and the Logic of Perception', Jaakko Hintikka suggests ways in which a model set theoretic semantics for perception might be formulated. Like its counterparts in the semantics for knowledge and belief introduced in Hintikka's Knowledge and Belief, the semantics for perception is based upon the concept of a model structure consisting of model sets, or "complete novels" as they're called in 'On the Logic of Perception'. These model sets are for Hintikka sets of sentences in a specified language whose membership is subject to conditions (set forth in Knowledge and Belief) designed in such a way that they may intuitively be regarded as descriptions of possible states of affairs.

A 'complete novel' is defined by Hintikka as "a set of sentences in some given language which is consistent but which cannot be enlarged without making it inconsistent", and a 'world' is "precisely what such a complete novel describes". At any given time the totality of complete novels belonging to a model structure may be divided into the set of those which are compatible with everything an agent perceives and those which are not compatible with everything the same agent perceives (a complete novel \( w \) is compatible with everything \( S \) perceives relative to some specified language if and only if for no \( p \) is it the case that (i) \( S \) perceives that \( p \), and (ii) \( \neg p \) is a member of \( w \)). And it is on the basis of this distinction that the door is open for 'perceives that' to be treated as a modal operator.

Assuming, for the sake of simplicity, that our concern throughout
is limited to what is perceived by a single agent Jones, the formula "Jones perceives that \( p \)" will belong to a complete novel \( w \) if and only if the formula \( p \) belongs to every complete novel compatible with everything Jones Perceives. And a formula is to be regarded as true in a world just in case it belongs to a complete novel which describes it. Therefore, we may think of a state of affairs as being perceived by Jones if and only if it holds in every possible world compatible with everything Jones perceives.

To vastly complicate matters, it is allowed (as in Hintikka's system of knowledge and belief) that variables occur free within the scope of the modal operator 'perceives that' and get bound from the outside. To put it another way, there are circumstances in which one is allowed to "quantify in" past the occurrence of one or more 'perceives that' operators. What these circumstances amount to are not specified in either 'On the Logic of Perception' or 'Information, Causality, and the Logic of Perception' (which we'll henceforth refer to as 'OLP' and 'ICLP', respectively), but they may be formulated on the basis of rules Hintikka gives in a number of different places for quantifying past knowledge operators. This we shall do shortly.

There are essentially two crucial differences between the semantics Hintikka develops for knowledge in *Knowledge and Belief* and the semantics for perception he sketches in the two perception papers. In the former, requirements are placed upon the compatibility relation between model sets: The relation must be reflexive and transitive.\(^3\) No such requirements are placed upon perceptually compatible worlds. Thus no success presuppositions are incorporated into the concept of perception.
with which the development of the formal semantics is concerned. One could easily incorporate reflexivity, but the resulting system would be substantially less interesting. "If we assume the success condition we cannot discuss such epistemically interesting problems as illusions, hallucinations, perceptual mistakes, impossible objects, etc."4

The second of the differences is of more fundamental importance, and it's here that we shall concentrate our attention. It is alleged that the logic of perception is really a bit more sophisticated than anything which can be captured in an ordinary semantics for quantified modal logic (as in Knowledge and Belief). Added machinery -- in the form of a second set of quantifiers -- is needed to capture the complexities inherent in the logic of perception. And so Hintikka's semantics for perception is equipped with a second set of quantifiers "(∃x)" and "(∀x)", in addition to the ordinary quantifiers "(Ex)" and "(x))."5

In this chapter we shall (i) Indicate how a formal semantics without the extra quantifiers may be characterized, (ii) Determine the role in Hintikka's system of the extra quantifiers, (iii) Argue that one of the major advantages Hintikka cites for the adoption of the new quantifiers -- giving rise to the formal expression of English "direct object" locutions -- is not an advantage gained by adding the new quantifiers, and (iv) Try to determine precisely the areas of increased expressibility afforded by the new quantifiers. * * *

We shall adopt the convention of designating the perceptual modal operator as 'Perceives'. Since we shall speak merely of what a
single agent (Jones) perceives, there will be no need to make use of a subscript identifying the agent in question. It might be thought that we are dangerously oversimplifying matters by concentrating upon what only a single agent perceives, but Hintikka points out (correctly, I believe) that it is a rather straightforward matter to generalize the system to provide for two or more agents.

A semantics of the sort Hintikka has in mind for a system involving a single agent and no epistemic operators other than 'Perceives' may be roughly characterized in the following way. Let a domain of complete novels be given together with a two-place relation $R$ taking complete novels as arguments. Assign to each complete novel a domain of individuals in such a way that these individuals are allowed membership in more than one domain (an individual $b$ can be said to belong to the domain of a complete novel if and only if $'(Ex) (b=x)'$ belongs to the novel). Let a member of the set of complete novels be chosen as the actual novel. And to complete the basic picture, choose appropriate free singular terms (constant terms), predicate terms, the identity sign, variables, and the usual connectives.

To this we add the model operator 'Perceives' defined in the manner already indicated and the standard quantifiers '$(Ex)' and '$(x)' defined in such a way as to rule out unrestricted Universal Instantiation and Existential Generalization.

In particular, let 'Perceives' be an operator such that 'Perceives $A$' is a member of a complete novel $w$ if and only if the formula '$A$' is a member of every complete novel $w'$ such that $Rw'$. Rules governing the use of the standard quantifiers go as follows:
Let 'A' be a well-formed formula and w any complete novel.

(i) Case One - The variable 'x' does not occur free inside an occurrence of 'Perceives'. Then if '(Ex)Ax' $\in$ w, 'A(b/x)' $\in$ w and '(Ex) (b=x)' $\in$ w, for some 'b'. And if '(x)Ax' $\in$ w, then if '(Ex)(b=x)' $\in$ w, then 'A(b/x)' $\in$ w.

(ii) Case Two - The variable 'x' occurs both within and without the scope of a single occurrence of 'Perceives'. Then if '(Ex)Ax' $\in$ w, 'A(b/x)' $\in$ w and '(Ex)(b=x & Perceives b=x)' $\in$ w, for some 'b'. And if '(x)Ax' $\in$ w, then if '(Ex) (b=x & Perceives b=x)' $\in$ w, 'A(b/x)' $\in$ w.

(iii) Case Three - The variable 'x' occurs only within the scope of a single occurrence of 'Perceives'. Then if '(Ex)Ax' $\in$ w, 'A(b/x)' $\in$ w and '(Ex)(Perceives b=x)' $\in$ w, for some 'b'. And if '(x)Ax' $\in$ w, then if '(Ex)(Perceives b=x) ' $\in$ w, 'A(b/x)' $\in$ w.

In addition we have the following rules:

(i) If '¬Perceives A' $\in$ w, then '¬A' $\in$ w*, for some w* such that R(w*)

(ii) If '(Ex)(b=x & Perceives b=x)' $\in$ w, then '(Ex) (b=x)' $\in$ w.

(iii) If 'A(b/x)' $\in$ w, and 'b' occurs free within the scope of 'Perceives', then if 'Perceives b=c' $\in$ w, 'A(c/x)' $\in$ w.

These rules may be easily generalized to handle formulas in which the variable bound by '(Ex)' or '(x)' occurs within the scope of multiple occurrences of 'Perceives'. For example, if '(Ex) (Perceives (Perceives Ax))' belongs to a complete novel w, we will be allowed to infer both 'Perceives(Perceives Ab))' $\in$ w and '(Ex)(Perceives(Perceives b=x))' $\in$ w, for some 'b'. An attempt is made by Hintikka in 'Existential
and Uniqueness Presuppositions' to codify the general requisite conditions for quantifying in any context containing any number of epistemic operators. For our purposes, however, it is sufficient to concentrate upon single modality contexts.

It may seem puzzling just why cases two and three above are given separate formal treatments. It may seem strange, for example, that '(Ex)(Perceives Ax)' and '(Ex)(x=x & Perceives Ax)' fail to be equivalent in the system we have just described. Why does Hintikka not treat all cases of quantifying in alike?

Hintikka's reasons for and defense of this particular formal distinction are quite complex and cannot be examined here in detail. Suffice it to say that this move makes it possible, Hintikka believes, to formally capture English "perceives who" constructions involving non-existent entities. We might, for example, wish to say of a man experiencing a hallucination of a person that he perceives who the person is without thereby implying that the person exists. By separating cases two and three in the manner indicated it is possible to distinguish between those instances in which perceiving who implies existence and those instances in which it does not.

Our account of Hintikka's semantics for perception must be accompanied by an important word of caution. It is alleged by Hintikka both in 'The Semantics of Modal Notions and the Indeterminacy of Ontology' and 'Objects of Knowledge and Belief' that difficulties arise in connection with the notion of "prefabricated possible individuals".

I am not convinced that the domain of possible individuals is anything we can start from in the sense of take for granted, at least not in some of the most
important philosophic applications of modal logic. I am not sure, either, that all the possible individuals we in some sense have to deal with can eventually be pooled into one big happy domain.

Suppose...that one begins by postulating a fixed supply of prefabricated individuals. Then one obtains a semantics which could function as an actual means of communication, it seems to us, only if one could assume that there are no problems in principle about re-identifying one's individuals as they occur in the several possible worlds we are considering. Once this presuppositions is made explicit, however, it is also seen at once how gratuitous it is for most philosophically interesting purposes.

My own attitude indicated there 'The Semantics of Modal Notions and the Indeterminacy of Ontology' can be summed up as a deep suspicion of those "prefabricated possible individuals" which have recently become so popular.8

Given these suspicions it may seem puzzling that each complete novel in Hintikka's semantics comes equipped from the outset with a domain of well-defined individuals. Does this mean that his system comes equipped with "prefabricated possible individuals"?

The answer is "no", and it is here that we must draw an important distinction between those individuals which inhabit domains of model sets, on the one hand, and functions which map model sets to these individuals on the other. These functions are referred to by Hintikka as "individuals", and it is this latter category which includes what Hintikka refers to as "possible individuals". For our own purposes, let us henceforth refer to individuals inhabiting domains of complete novels as "individuals1" and functions from model sets to individuals1 as "individuals2" whenever there is apt to be any confusion. We can then say that Hintikka has in mind individuals2 and not individuals1 when he expresses suspicion of "prefabricated possible individuals". Thus, by introducing a
model structure, part of which involves domains of prefabricated individuals, we do not thereby introduce prefabricated individuals.

One final word before moving on. It seems quite clear that to tell whether a world is compatible with another all that needs to be done is to look for logical consistency between two sets of formulas. A world \( w \) is compatible with (or an alternative to) a world \( w' \) \((Rw'w)\) if and only if the conjunction of the members of the complete novel describing \( w \) is logically consistent with the conjunction of the set of formulas \( 'A' \) such that \( 'Perceives A' \) is true in \( w' \). This gives us the hoped for results that no world \( w \) is compatible with \( w' \) if what is perceived in \( w' \) happens to entail a contradiction and that every world is compatible with \( w' \) if the agent's only perceptions in \( w' \) are tautologies.

Nonetheless, Hintikka expresses a reluctance to accept this account:

It is to be noted that the notion "compatible with what a perceives" is to be taken as unanalyzable. To know what someone, say \( a \), perceives in a world \( w \) is to know what-is-compatible-with-what-he-perceives among other possible worlds \( w' \), and the latter notion -- it is a relation between the possible worlds \( w' \) and \( w \) -- turns out to be the more powerful one for the purposes of semantical analysis. Attempts to analyze it have turned up nothing useful. If one can list all the facts that \( a \) perceives in \( w \) one can of course define the perceptual alternatives to \( w \) as those worlds which are logically compatible with those lists in the sense in which all the members of the list are true in the world in question. But this does not accomplish anything new, and it rules out those perceptual situations in which we cannot specify in some particular language all the facts \( a \) perceives. In short, it does not allow for unverbalized perceptions.

Be that as it may, nothing in the present study turns, I believe on the distinction between verbalized and unverbalized perceptions. So
for purposes of simplicity we assume that our agent Jones is eloquent enough to verbalize all he perceives. This will then insure that the notion of compatibility we make use of is precisely the one described above.

* * *

We move now to consider Hintikka's introduction of a second pair of quantifiers.

The great interest of perceptual concepts for a philosopher of logic is due precisely to the fact that we all as a matter of fact use two different methods of individuation. One of them is the method of physical individuation indicated above, but the other is essentially different from it...When presented with descriptions of two different states of affairs compatible with what S sees, and with two different individuals figuring in these two respective descriptions, we ask whether they are identical as far as S's visual impressions are concerned, and often we can answer this question. The question therefore gives us another method of individuating objects in contexts in which we are talking of what someone sees at a given moment of time...we shall call individuals so cross-identified "perceptually individuated objects".

There may, for instance, be a man in front of a of whom a does not see who he is. Then descriptively speaking the man in front of a is a different (descriptively individuated) person in different possible worlds compatible with everything a sees. But in each such world there will have to be a man in front of a ...Obviously we can in principle use this fact for the purpose of cross-identification and as though it were a trans world heir line through all these men in their respective worlds.

The introduction of the new quantifiers is now quite simple. Having introduced these new perceptually individuated objects (called "perceptual individuals" or "perceptual objects" for short) Hintikka permits variables to range over them. And in order to bind these variables special new "perceptual quantifiers" are added to the language.
What is the nature of these perceptual individuals? The clearest indication Hintikka gives us is that they can be envisaged as functions which from each world over which they're defined pick out one individual from its domain (or less formally described in ICLP as world lines drawn through individuals in the different compatible worlds; notice, then, that perceptual individuals are individuals). These individuals so picked out are said to be the "same" as far as the agent in question's perceptions are concerned. Suppose, for example, that Jones perceives that the man in front of him is bald. Suppose further that there is no man of whom Jones perceives that he is the man in front of him. Then the location 'the man in front of him' picks out different physically individuated objects (called 'physical objects' for short) in different worlds compatible with everything Jones perceives.

We may suppose, for example, that it's compatible with everything Jones perceives that the man in front of him is Y.A. Tittle. Then in at least one compatible world Y.A. Tittle really is the man; and perhaps in another world Telly Savales is the man in front of Jones. "Because of this," Hintikka urges, "we may say that from the point of view of Jones's perceptual situation they are after all one and the same man -- the man in front of him." 12

Physical individuals, by contrast, are described by Hintikka as those functions whose value at each complete novel w coincides with the reference at w of some constant term 'b' which is such that (at the actual world) Jones perceptually recognizes b or perceives who b is. Therefore, if Jones perceives who the man in front of him is, the term 'the man in front of him' picks out the same
physically individuated object in each world compatible with everything Jones perceives. This means that we now have two different criteria for determining trans-world identity. If two individuals, x and y, both turn up in the range of a physically individuated object they are the "same" in one sense, and if both turn up in the range of a perceptually individuated object they are the "same" in another sense.

Whereas physical and perceptual individuals are both to be thought of as functions which map worlds to members of their respective domains, it will be intuitively helpful, I believe, to think of physical individuals as constant functions and perceptual individuals as (in general) non-constant functions. Alternatively, we may (in the spirit of ICLP) think of them as "rigid" and "wobbly" world lines, respectively. There is, of course, nothing in the formal semantics which gives us any more reason to conceive of physical individuals as constant functions than to conceive of perceptual individuals as constant functions. But we shall establish this convention for the purposes of our own discussion to aid us in better picturing the way in which the two kinds of quantifiers relate to one another. To be explicit, if the value of a physical individual at a world w is an individual x, then x is the value of the physical individual at every world compatible with w. From this, of course, it does not follow that every constant function defined over a set of worlds compatible with w is a physical individual.

Occasionally perceptual individuals do, the way we are now viewing the situation, behave as constant functions, and this phenomenon occurs precisely in those cases where there is a perceptual individual
b such that Jones happens to perceive who b is. If Jones correctly perceives of the man in front of him that he is Y.A. Tittle, it is no longer compatible with everything Jones perceives that Savales or anyone else is the man in front of him. The perceptual individual now picks Tittle from every compatible world.

Suppose we let 'm' be short for 'the man in front of him (Jones)'. Then to say that 'm' designates a perceptual individual is (formally speaking) to say that '($\exists x$)(Perceives m=x)' is the case; to say that 'm' designates a physical individual is to say that '($\exists x$)(Perceives m=x)' is the case ("Jones perceives who m is"). Therefore, '($\exists x$)$\forall y$(Perceives x=y & m=x)' is true in precisely those cases where 'm' picks out a perceptual individual which happens to coincide with a physical individual. And by so coinciding the perceptual individual is forced to become a constant function; in the case at hand, it is now forced to choose Tittle and no one else at each compatible world.

We turn now to the task of finding a place for perceptual individuals in the model set theoretic semantics sketched above. In this regard Hintikka is of very little help. Nowhere does he speak of rules governing the use of the new quantifiers. Nor does he give any indication as to how we might begin to decide just which functions definable over worlds count as perceptual individuals and which do not. Nevertheless, we are given a couple of important clues relating to the behavior of perceptual individuals.

First, every individual chosen by a perceptual individual at any world w must belong to the domain of w; no individual failing to exist in a given world may be the assignment of a perceptual individual
at that world. If an agent $m$ perceives that there is a man in front of him but does not perceive who he is, then, says Hintikka, "The man in front of him (let us call him $m$) is a different individual (different person) in some of the relevant possible states of affairs... In all of these different states of affairs, however, there has to be a man in front of him. (Otherwise the state of affairs in question would not be compatible with what $m$ sees)."

13 Hintikka makes the same point in ICLP in a passage we have already quoted (page 9 above).

Second, there is a need to restrict the domains over which perceptual individuals are defined; they cannot be defined arbitrarily over members of $W$. For suppose that no limits at all are placed upon the worlds which may qualify as members of domains of perceptual individuals. Then suppose that each of two perceptual individuals map each compatible world to exactly the same individuals. But suppose that the mappings diverge when they come to some world not compatible with everything the agent in question perceives. Then we are forced to admit that the perceptual individuals are distinct, even though they seem to fulfill every conceivable requirement for being considered the same perceptual individual. An example will probably make this clearer.

Suppose that Jones perceives that there is a man in front of him, that there is a man aiming a rifle at him, and that the man in front of him is identical to the man aiming a rifle at him. Suppose further that he fails to perceive just who this man is. It seems clear in this example that the perceptual individual designated by 'the man in front of him' and the perceptual individual designated by
'the man aiming a rifle at him' should turn out to be identical. And it seems equally clear that their identity should not be called into question by the possibility that the perceptual functions extend to a non-compatible world where, as it turns out, the man aiming the rifle is distinct from the man in front of Jones. To avoid all this it seems quite necessary to restrict perceptual individuals to worlds compatible with what is being perceived.

Hintikka seems to have been aware of the need for this restriction:

By cross-identification we of course mean here telling which individual in one possible world is identical with which individual in another (identification across the boundaries of possible worlds). The possible worlds involved here are of course those compatible with what a perceives.\(^4\)

Yet we are going to find that Hintikka allows for one exception to this restriction: Under certain conditions the perceptual individual under consideration may be defined at the actual world even when it is not an alternative to itself. We return to this matter later. For the time being we shall assume that perceptual individuals as a general rule never extend beyond the set of compatible worlds.

* * *

Given these clues relating to the nature of perceptual individuals, it may seem pertinent now to try to specify the formal conditions which suffice for the formation of a perceptual individual. Under what circumstances does a function defined over the set of compatible worlds count as a perceptual individual? At first approximation it may seem sufficient that there exist a function which (i) Assigns to each compatible world some member of its domain, and (ii) Is undefined elsewhere. However, there are a great many bizarre functions which
must surely be ruled out as perceptual individuals, functions which manage to pick out entirely random objects from the various compatible worlds. Some restriction must be placed upon these functions to choose those suitable to qualify as perceptual individuals. And some criterion must be made to make that choice.

Unfortunately, I know of no simple way in which such a criterion may be spelled out, and I shall not attempt to arrive at one. Nevertheless, I shall argue in this section that a formal semantical characterization of perceptual individuals is not required in order to formulate truth-conditions for at least one important subclass of formulas containing perceptual quantifiers. Thus, we shall bypass entirely the notion of a perceptual individual for the time being.

So far we have said nothing about why a second set of quantifiers is required in Hintikka's system. Presumably there are English sentences (describing Jones's perceptions) which can be formally captured in Hintikka's system only by making use of special quantifiers which in effect range over intensions. If so, it is certainly a question of major interest to determine which sorts of sentences these might be. Surprisingly, Hintikka does not have much to say about the increase in expressive powers which results by employing perceptual quantifiers; with one exception, he almost completely ignores the issue.

That one exception, however, seems to be regarded by Hintikka as quite crucial. It is the set of English sentences which contain what he calls "direct-object constructions".

Perceptual cross-identification is presupposed in the truth-conditions of such direct-object constructions as 'a sees b'. The point is perhaps explained most quickly by pointing out that for a to
see b (direct object construction) is for b to find a place among a's visual objects, that is to say, among the individuals which a can locate in his visual space. A simple argument shows that this is the case when we have

\[(3) (\exists x)(a \text{ sees that } (b=x))\]

with a quantifier '3' relying on perceptual cross-identification.\(^{15}\)

Here the appearance of perceptual quantifiers is used to give expression to locutions (in English) of the form 'a perceives b' or 'a sees b', where 'b' is a singular term. To say "Jones perceives Smith" is simply to say that (for Jones) Smith is a perceptual individual, and we can easily translate this sentence into the language of Hintikka's system as '(\exists x)(\text{Perceives Smith}=x)'. And so it is that "we have now found an analysis of the direct-object constructions in terms of quantifiers and of 'Perceives that'".\(^{16}\) As Clark observes, "Hintikka exploits the two types of quantifiers to give expression to direct-object constructions".\(^{17}\)

It is Hintikka's claim, therefore, that it is the addition of perceptual quantifiers to his system which makes it possible to formally render English direct-object constructions. Suppose we investigate this claim by turning to consider formulas of the form '(\exists x)(\text{Perceives } Ax)'. Assuming we have made sense of what it means to say that an open formula of one variable is 'satisfied' by an individually in a world, it is clear that '(\exists x) (\text{Perceives } Ax)' is true in a world w if and only if there is a perceptual individual which assigns to every world compatible with w a member of its domain in such a way that 'Ax' is satisfied by that member in that world.

Given this characterization of '(\exists x)(\text{Perceives } Ax)' it should be clear that '(\exists x)(\text{Perceives } Ax)' virtually implies 'Perceives (Ex)
If there is a perceptual individual making assignments in each compatible world, the members of its range must show up in the various worlds. For suppose that 'Perceives (Ex) (Ax)' is false in w. Then in some world w' compatible with w ' (Ex) (Ax)' is false. So nothing in the domain of w' satisfies 'Ax' in w'. Hence there can be no function assigning each compatible world an element from its domain satisfying 'Ax'. In particular, no perceptual individual can do this. So '(3x)(Perceives Ax)' is false in w.

And this should seem reasonable. If a man perceives an apple, then it ought to be the case that there is at least one apple to be found in every world compatible with everything he perceives. It is not compatible with everything he perceives that there are no apples if he really does perceive an apple. Therefore, as long as we specify that perceptual individuals map each world to an individual which exists in that world, it is quite easy to see that '(3x)(Perceives Ax)' virtually implies 'Perceives (Ex)(Ax)'.

Having shown that 'Perceives (Ex)(Ax)' is a necessary condition for the truth of '(3x)(Perceives Ax)', the question arises whether 'Perceives (Ex)(Ax)' is at the same time a sufficient condition. Does the existence of apples in every single world compatible with everything Jones perceives enable us to conclude that Jones perceives an apple? Again, the answer seems to be "yes". And this can be shown as follows.

Suppose, first of all, that '(3x)(Perceives Ax)' is false in w. We shall argue that 'Perceives (Ex)(Ax)' is then false. If '(3x)(Perceives Ax)' is false in w, then our agent (Jones) fails to perceive that something satisfies 'Ax' in w. So by W-completeness (see
footnote 12) there is some world $w'$, compatible with everything Jones perceives, such that 'Ax' fails to be satisfied by something which exists in $w'$. Hence $-\exists x(\exists y)$ is true in $w'$. So it is not the case that $\exists x(\exists y)$ is true in every world compatible with $w$. But if not, then by the way we set up truth-conditions for 'Perceives', $-\exists x(\exists y)$ is true in $w$; and so $\exists x(\exists y)$ is false in $w$, which was to be shown.

Intuitively, the idea is this. If something red shows up in every compatible world, there must be a red perceptual object. For if nothing in Jones's perceptual field appears red, then it's compatible with what he perceives that nothing at all is red, and so there must sooner or later come a compatible world in which nothing at all is red. Thus, the presence of red things -- whatever they are -- in each compatible world insures the presence of a red perceptual individual. We have no idea where the perceptual world line goes from world to world, but we know there must be one. And so truth conditions for formulas of the form $(\exists x)(Perceives Ax)$ may be given independently of specifying necessary and sufficient conditions for a given function's qualifying as a perceptual individual.

In short, we have shown that $(\exists x)$ in its most basic role (binding variable occurring inside an epistemic operator) may be characterized entirely in terms of the ordinary quantifier $(\exists y)$.

A formula of the form $(\exists x)(Perceives Ax)$ is true in a world $w$ if and only if $(\exists y)(Ax)$ is true in every world $w'$ such that $Rww'$. A man perceives an apple if and only if there exists at least one apple in every compatible world.
It now looks as though the employment of Hintikka's perceptual quantifiers is quite unnecessary to formally capture English direct-object locutions. If sentences of the form "Jones perceives b" are to be symbolized by formulas of the form '(∃x)(Perceives b=x)', and if formulas of the form '(∃x)(Perceives Ax)' are virtually equivalent to formulas of the form 'Perceives (Ex)(Ax)', then sentences of the form "Jones perceives b" may equally well be symbolized by formulas in which no perceptual quantifiers occur. It may well be the case that '(∃x)(Perceives b=x)' in some way gives us more insight into the logic underlying English direct-object constructions than something like 'Perceives(Ex)(b=x)'. But the point of the matter is that, given everything we have shown, perceptual quantifiers are not required in the formal expression of English direct-object constructions. If there are English sentences whose formal translation requires perceptual quantifiers, these sentences require the quantifiers for reasons apart from the need to capture direct-object constructions.

We now move to consider some objections which might arise in connection with our argument that '(∃x)(Perceives Ax)' and 'Perceives (Ex)(Ax)' are virtually equivalent. From an ordinary philosophical point of view there appear to be counter-examples to such an equivalence. Suppose, in the first place, that Jones perceives a dagger, and suppose that the dagger is a perceptual object. Then '(∃x) (Perceives(x is a dagger))' is true. But suppose Jones is hallucinating and that there really is no dagger. Suppose, moreover, that Jones realizes full well that he's hallucinating and that no dagger is present. Given these circumstances it seems a mistake to
conclude that Jones perceives that a dagger exists. That is, 'Perceives(Ex)(x is a dagger)' seems false.

The appropriate response here is that if 'Perceives (Ex)(x is a dagger)' seems false, then more is being read into the concept of perceiving than is warranted. We have been operating throughout with an especially weak notion of perceiving, and to suppose that 'Perceives A' and 'Knows not-A' are inconsistent is at this point unjustified. These are inconsistent if and only if we are in general guaranteed some state of affairs compatible both with what is perceived and with what is known. By making R reflexive we guarantee such a mutually compatible state of affairs (the actual world), but in the system under consideration R is not reflexive. Hence there's no reason to conclude that just because Jones knows there's no dagger out there, he doesn't perceive there is one.

A second counter-example to the equivalence involves the inference going in the opposite direction. Imagine a situation where Jones is observing the gasoline gauge in his automobile. By noticing that the gauge indicates a half-full tank Jones perceives that there is gasoline in his tank. Yet from 'Perceives(Ex)(x is gas)' it is supposed to follow that '(∃x)(Perceives(x is gas))' is true. But '(∃x)(Perceives(x is gas))' is certainly not true; at no time does Jones ever perceive gasoline. He perceives that there is gasoline without ever perceiving gasoline. And so the inference does not seem valid after all.

The trouble here comes with interpreting the English sentence "Jones perceives that there is gasoline". It is assumed in the example that 'Perceives(Ex)(x is gas))' correctly captures this sentence, but
a closer examination of the facts reveals that this is not so. 'Perceives(Ex)(x is gas)' is true just in case there is gasoline in every world compatible with what Jones perceives. But although there is a fuel gauge in every world, there are worlds in which no gasoline is present in his tank (otherwise he'd be perceiving gasoline directly). Strictly speaking, it is only inferred from what he perceives that gasoline is present, and since '(Ex)(x is gas)' is not a logical consequence of what is perceived, there are complete novels compatible with what is perceived of which '¬(Ex)(x is gas)' is a member.

Perhaps the real motivation behind this counter-example is the desire for a modified notion of 'compatibility', according to which that which may be "reasonably inferred" from what is perceived is true in each compatible world. As Kant observes,

Since we have constantly to make use of inference, and so end by being completely accustomed to it, we frequently... treat as being immediately perceived what has really only been inferred.20

Such a modification might indeed produce a semantics more in line with our common usage of "perceives", but its implementation depends upon a precise account of what English speakers would consider a reasonable inference. And such an account is certainly not obvious.21

It may be objected, finally, that under our interpretation of '(∃x)' the formula '(∃x)(Perceives ¬(Ey)(x=y))' turns out to be self-contradictory (indefensible). It ascribes to Jones an inconsistent perception: that something both does and does not exist. Yet surely Jones is able to perceive that something in his visual field fails to exist; he may perceive a dagger and, observing that it appears transparent and wave-like, conclude that it's not really present.
To this we reply that Jones surely can perceive that something in his visual field fails to exist; but when he does he simply perceives that something he perceives fails to exist. He perceives that something fails to exist, and this something is just one of the things he perceives. Therefore, one of the things perceived by Jones is perceived by Jones not to exist. And this is easily symbolized as 'Perceives-(Ey)((Ex)(x=y) & Perceives (Ex)(x=y))', a formula which, unlike (∃x)(Perceives -(Ey)(x=y))', is perfectly consistent.

So far we have argued that formulas of the form ' (∃x)(Perceives b=x)' are (i) Employed by Hintikka to capture English direct-object constructions, and (ii) Can be reduced to formulas which do not contain occurrences of perceptual quantifiers. From this we cannot, however, conclude that all formulas employed by Hintikka to capture direct-object constructions can be reduced to formulas lacking perceptual quantifiers without knowing in addition that (iii) Formulas of the form ' (∃x)(Perceives b=x)' are the only formulas employed by Hintikka to capture direct-object constructions. And we shall now discover surprisingly that (iii) is actually false.

The truth of the matter is that Hintikka makes a distinction in both of his papers between two senses of the English sentence "Jones perceives b". In OLP the two senses are described as follows: According to one sense of "Jones perceives b" Jones perceives "the individual in question", while according to the second Jones perceives b "whoever he is or may be"; the first is symbolized by Hintikka as ' (∃x)(Perceives b=x)' and the second as ' (∃x)(b=x & Perceives (Ey) (x=y))'. Suppose we let 'b' be short for 'the man on the horizon carrying a flag'. Then by saying "Jones perceives b" we may be
saying either of two things. First, we may be saying that Jones perceives that there's something on the horizon, that it's a man, and that he's carrying a flag; this is the stronger sense of "Jones perceives b" and is expressed by \((\exists x)(\text{Perceives } b=x)\). But according to the weaker sense of "Jones perceives b", Jones perceives only that the individual, who in fact is b, exists. He perceives something but fails to perceive either that it's a man, it's on the horizon, or that it's carrying a flag; this weaker sense is expressed by \((\exists x)(b=x \& \text{Perceives}(E_y)(x=y))\). He perceives b but fails to perceive that what he perceives is b. In ICLP Hintikka refers to this distinction as a "contrast between statements de dicto and de re"; on the de dicto reading of "Jones perceives b" Jones perceives that what he perceives is in fact b, while on the de re reading Jones simply perceives an individual which happens to be b.

The upshot of this distinction is that, while Hintikka employs formulas of the form \((\exists x)(\text{Perceives } b=x)\) to give formal expression to sentences of the form "Jones perceives b", it is not true that formulas of the form \((\exists x)(\text{Perceives } b=x)\) are the only formulas employed in this capacity. In particular, the de re reading of "Jones perceives b" is captured instead by a formula in which variables bound by a perceptual quantifier occur both within and without the scope of the epistemic operator. And so far we have shown nothing to the effect that formulas such as these can be reduced to formulas which contain only ordinary quantifiers. Therefore, we have not really shown at all that perceptual quantifiers are not required to give formal expression to sentences of the form "Jones perceives b". We have shown that perceptual quantifiers are not required to express
formally direct-object constructions, provided they are given a de dicto reading. It may still well be the case that, given a de re reading, sentences containing such constructions really do require the powerful resources of quantifiers ranging over intensions.

I shall now state quite categorically that, so far as I can see, formulas of the form '(∃x)(Ax & Perceives Bx)' really cannot in general be reduced to formulas in which perceptual quantifiers are eliminated. Therefore, the de re sense of sentences of the form "Jones perceives b" does indeed seem to require the use of perceptual quantifiers in order to be formally expressed. Nevertheless, in the remainder of this section I shall argue that formulas of the form '(∃x)(Ax & Perceives Bx)' turn out on Hintikka's approach to bring with them certain difficulties of interpretation; they turn out to defy any sort of clear interpretation (relative to possible worlds) required to understand what they assert. Hence we shall conclude that the formal expression of direct-object constructions is not an advantage gained by the addition of perceptual quantifiers. On one reading of such constructions the quantifiers appear unnecessary, and on the other reading the quantifiers are employed in a way that fails to make clear sense relative to a possible worlds interpretation of the formulas in which they appear.

We begin by asking precisely what formulas of the form '(∃x)(Ax & Perceives Bx)' are supposed to assert (where we assume there are no occurrences of 'Perceives' in 'A'). The answer, presumably, is that '(∃x)(Ax & Perceives Bx)' asserts that there is a perceptual individual which (i) Picks out of each compatible world some member of that world's domain which satisfies 'Bx' in that world, and (ii) Picks
from the actual world something satisfying 'Ax' in the actual world. The formula asserts that, in other words, something perceived by Jones to be B is as a matter of fact A. What could possibly be problematic in interpreting a formula such as this one?

The problem arises entirely out of the failure of Hintikka's complete novels to be governed by reflexivity. Since it is not always the case that the actual world numbers among its own alternatives (that is, is not always compatible with itself), it turns out that a perceptual individual making assignments in the manner required by '(∃x)(Ax & Perceives Bx)' forces us to violate the rule that perceptual individuals never extend further than the set of compatible worlds. Of course, there is nothing sacred about this rule; but by extending itself to the actual world even when it fails to be one of the compatible worlds, the perceptual individual is forced to choose something in the actual world. And this does cause problems; for when the actual world is not one of its own alternatives it is often extremely difficult to see what, if anything, this something is to be.

Consider the sentence, "What Jones perceives to be a fat man in front of him is really not a man at all." Letting '(∃x)((Perceives (FMx & Fx)) & ~Mx)' be the formal expression of this sentence, it is clear that we are face to face with a difficult puzzle. It is easy to trace the perceptual world line through the set of compatible worlds; we simply pick up whoever happens at each world to be the fat man in front of Jones. But at the actual world there is no such man, and yet we are required to choose something. What, therefore, are we to choose? And how can we draw up a criterion according to which we
can in any given situation decide which element of the actual world is the value of the perceptual individual at the actual world?

Hintikka realizes this difficulty and offers the following solution. Suppose we let \( w \) be the actual world.

How then do we cross-identify between \( w \) and its alternatives? This question is crucial for the truth-conditions for such propositions as \( (\exists x)(x=d \land \text{Perceives } x=c) \).

The trans world heir lines which are (go together with) values of '\( x \)' here must be extended to the actual world in order for us to be able to say, as we are attempting to do in the above formula, that in the actual world one of them is (picks out) \( d \). From this we can see what the problem is intuitively. The question concerns the principles according to which we say that one of \( a \)'s perceptual objects is or is not identical with an object in the actual world... The 'trans world heir lines' connecting these several worlds are as it were drawn differently when it comes to connect the actual world with its perceptual alternatives than when it is required to weave together these alternatives. Since these world lines are involved essentially with the truth-conditions of quantified statements, it is seen that the truth-conditions of statements in which one quantifies both into a perceptual context and outside it (thus requiring the alternatives to be compared with the actual world) involve considerations essentially disparate from those involved in the truth-conditions for statements in which one merely quantifies in. The former turn also on causal considerations... Grice... registers some of the difficulties in spelling out precisely what the causal connection is which has to obtain between one of \( a \)'s perceptual objects and an actual object in his environment before we can tie them together with the same world line. It seems to me that the problem of giving a precise characterization of this connection is due more to general difficulties in analyzing causal notions than to the special features of causal connections relied on in perception.\(^{23}\)

Consider once again the sentence, "What Jones perceives to be a fat man in front of him is not a man at all". The problem is to determine which object in the actual world is picked out be the perceptual individual, there being no fat man in front of Jones in the actual
world. Hintikka's solution is to determine what exactly in $w$ caused Jones to perceive a fat man and simply tie the loose end of the world line to it. If it's a large stump, then the world line extends to the stump. And surely there is something reasonable about this; if a stump really caused Jones to perceive a fat man, then we'd be inclined to say that what he perceives is really not a fat man but rather a stump. In other words, \((\exists x)((\text{Perceives}(Fx, x) \& Fx) \& x \text{ is a stump})\) seems agreeable. The stump is part of the same world line that picks out fat men in each compatible world. If nothing more, all of this accords remarkably well with our ordinary perceptual talk.

Hintikka's solution is intuitively appealing, but how might all of this be captured in the formal semantics? How can we analyze this causal connection needed to join the actual world to its alternatives? Hintikka certainly leaves us in the dark here; he speaks of the "difficulties in analyzing causal notions" and seems to regard them lightly as though an exact characterization of a semantics for 'Perceives' is no longer his concern.

Yet there really seems to be no way to take Hintikka's solution seriously until we can analyze the notion, 'x causes y to perceive that p' in terms formally specifiable in Hintikka's system. Obviously, the prospects of cooking up such an analysis are not encouraging. But notice that even if we are able to find an analysis of 'x causes y to perceive that p' we are still hard pressed to account for hallucinations. There are times, it would seem, that no actually existing object can be cited as the cause for Jones's perceiving certain things to be the case (excepting, of course, indirect causes such as alcohol). Thus, the sentence "What Jones perceives to be a mountain lion is
actually nothing at all" would resist the type of treatment Hintikka offers.

By all indications, therefore, it seems that Hintikka's system (by virtue of lacking reflexivity) is incapable of providing a satisfactory, intuitive possible worlds interpretation for formulas of the form '($\exists x$)(Ax & Perceives Bx)'. There simply are no clear guidelines for deciding the value of the perceptual individual at the actual world. Permitting perceptual individuals to extend beyond the safety of the set of compatible worlds creates nothing but chaos in the formal semantics.

We conclude that it is not at all clear what, with a particular possible worlds interpretation in mind, formulas of the form '($\exists x$)(Ax & Perceives Bx)' are supposed to assert. While we can develop clear intuitions as to what is asserted by '($\exists x$)(Perceives Ax)' by studying its truth-conditions relative to a possible worlds interpretation, we can do nothing analogous in the case of '($\exists x$)(Ax & Perceives Bx)'. Thus, while there is potentially much to be gained by formally rendering certain English constructions by formulas of the form '($\exists x$)(Perceives Ax)', there seems to be little or no point in doing the same with '($\exists x$)(Ax & Perceives Bx)' as long as reflexivity fails to hold. De re direct object constructions may not be adequately handled in a system without perceptual quantifiers, but there is certainly nothing to be gained by dealing with them in a system with the extra quantifiers in such a way that a perceptual quantifier is forced to bind variables occurring both within and without the scope of the epistemic operator.

The difficulties with '($\exists x$)(Ax & Perceives Bx)', incidentally,
do not arise with formulas of the form \((\exists x)(Ax \& \text{Perceives } Bx)\). With \((\exists x)(Ax \& \text{Perceives } Bx)\) we are guaranteed the existence in the actual world of the individual picked out by the physical individual at the actual world (something we are not guaranteed with \((\exists x)(\text{Perceives } Ax)\), as we pointed out earlier). Thus, if there is an actual flesh-and-blood person with respect to whom Jones perceives that he B's, and if as a matter of fact this person A's, then no matter what we fill in for 'A' or 'B', still we are guaranteed someone in the actual world to which we tie the loose end of the world line, reflexivity or no reflexivity. We can take hold of this flesh-and-blood person and say of him that he is the person of whom Jones perceives that he B's. And we have no trouble finding this flesh-and-blood person or mistaking who this person is. By the very nature of physical individuation Jones perceives who he is, i.e., which individual in the domain of the actual world he is. We do not have to fuss with finding whatever physical thing (if any) happened to cause Jones's various perceptions.

**

We discovered earlier that an important subclass of formulas containing perceptual quantifiers -- those where \((\exists x)\) binds variables occurring strictly within the scope of 'Perceives' -- can be reduced to formulas lacking perceptual quantifiers. A reduction of this sort naturally raises the question whether further reductions of perceptually quantified formulas may be made. Most exciting of all perhaps is the question whether perceptual quantifiers can be done away with altogether in favor of ordinary quantifiers alone, and whether everything that can be said in the language of Hintikka's system
could be said without employing the extra quantifiers. In the final section we shall explore the question of formal expressibility and point to some specific instances where the new quantifiers really do seem to add to what can be said in Hintikka's system.

In his paper 'Reply' Romane Clark has a good deal to say about what he calls "Hintikka's double quantifier theory". It is ostensibly Clark's view that the perceptual quantifiers can be altogether eliminated from the system; he writes that the standard quantifier is "apparently adequate to the distinctions Hintikka wishes to draw" and goes on to suggest in rough fashion how in a certain number of instances this is the case. It is my belief, however, that were Clark to attempt to offer a completely general account, he would discover that there really are distinctions which can be made only by means of employing two sets of quantifiers.

Consider first the set of formulas in which a perceptual quantifier is directly followed by an ordinary quantifier of the opposite sort ("...(*x)(y)..." and "...(*x)(Ey)..."), both binding variables occurring within (but not without) the scope of 'Perceives'. We take this set to include all formulas equivalent to "...(*x)(y)..." and "...(*x)(Ey)..." constructions ("...(*x)-(Ey)...", etc.). Let us refer to formulas belonging to this set as "P-formulas". It shall then be my first contention that P-formulas cannot be reduced to formulas which contain no occurrences of perceptual quantifiers.

Suppose Jones perceives a number of people standing under a bridge. Each of the people is such that Jones perceives who he is, but suppose he fails to perceive what it is they are standing under. All he perceives is that a group of persons, each of which he recognizes,
is standing under a structure of some sort. Then there is a perceptual individual \( x \) such that for any physical individual \( y \) which picks up a person at the actual world, Jones perceives that \( x \) is over \( y \). This we symbolize by the P-formula '\((\exists x)(y)(y \text{ is a person Perceives}(x \text{ is over } y))'" (for simplicity let's just consider '\((\exists x)(y)(\text{Perceives}(x \text{ is over } y))'\).

It is my belief that '\((\exists x)(y)(\text{Perceives}(x \text{ is over } y))' cannot be rewritten in terms of ordinary quantifiers. The formula '\((\exists x)(y)(\text{Perceives}(x \text{ is over } y))'\), for example, makes the bridge a physical individual which by hypothesis it is not; it says, in other words, that Jones knows what it is the people are standing under. This particular difficulty may be remedied by trying something like '\((y)(\text{Perceives}(\exists x)(x \text{ is over } y))'\), but this formula allows for the possibility that over each person there appears a structure but that there are different structures appearing over different people. Finally, 'Perceives(\exists x)(y)(x \text{ is over } y)' fails to report that each person in the crowd is perceptually recognized by Jones.

The problem, crudely put, is this. In order to guarantee that persons in Jones's perceptual field are physically individuated objects, the '\((y)' must remain outside the modal operator and bind free variables occurring within. And to guarantee that the bridge is not a physically individuated object, the 'Ex' must occur inside the modal operator. But to preserve the sense of the original formula it is necessary that 'Ex' precede '\((y)'\. It's hard to see how all of this can be done at once, and so quantifying over perceptual individuals does seem unavoidable to capture what needs to be said here. The perceptual quantifier seems to get trapped outside ordinary
quantifiers in cases such as these, and once they are trapped in this way they cannot be eliminated. The situation is the same in the case of all P-formulas.

Consider next the class of formulas in which a universal perceptual quantifier binds variables occurring within but not without the scope of 'Perceives'. Suppose we refer to this class as the class of "Q-formulas" (and include '...(∃x)...' constructions, etc.). It shall then be my second claim that Q-formulas too contain perceptual quantifiers which cannot be translated out.27

Suppose that Jones visits a brickworks and that everything on which he lays eyes is covered with orange dust. Then to say in the language of Hintikka's system the English sentence, "Everything perceived by Jones is perceived by Jones to be orange", we simply employ the Q-formula '(∀x)(Perceives(x is orange))'. Now it may seem like an easy move from '(∀x)(Perceives(x is orange))' to the formula 'Perceives(x)(x is orange)' in the same way we argued earlier that '(∃x)(Perceives Ax)' can be reduced to 'Perceives(Ex) (Ax)'. Yet this is certainly not the case.

According to '(∀x)(Perceives(x is orange))', every perceptual individual assigns to every compatible world w something in the domain of w which satisfies 'x is orange' (in w). But although 'Perceives(x)(x is orange)' also guarantees that every member of every compatible world w chosen by a perceptual individual satisfies 'x is orange' in w, it says more. It guarantees in addition that those members of any compatible world w which are not chosen by any perceptual individual also satisfy 'x is orange' in w. Thus, 'Perceives(x)(x is orange)' virtually implies '(∀x)(Perceives(x is orange))', but the
reverse does not hold. Everything Jones perceives he perceives to be orange, but Jones certainly does not perceive that everything is orange; he does not perceive, for example, that his parents (back home in Ohio) are orange. It's compatible with everything he presently perceives that his parents are not orange. And so 'Perceives(x)(x is orange))' is much to strong. We want only to assert that the things he perceives are perceived to be orange.

The remedy seems clear. Find an antecedent 'Ax' to insert within the scope of '(x)' in 'Perceives(x)(x is orange))' (yielding 'Perceives(x)(Ax→x is orange))' which simply asserts that Jones perceives x. Then instead of saying that Jones perceives that everything is orange we'll be saying that Jones perceives that only a certain number of things are orange, namely the things he perceives, and this is presumably what we want. To follow through on this suggestion we might try something like 'Perceives(x)(Perceives (x=x)→x is orange)' or else 'Perceives(x)(Perceives(Ey)(x=y)→x is orange)'. Will either of these do the trick?

Unfortunately not. For we now have a situation where an ordinary quantifier binds variables occurring within the scope of an occurrence of 'Perceives', and this means that we are suddenly now dealing with physical individuals. 'Perceives(x)(Perceives(x=x)→x is orange)' does not say that Jones perceives that everything he perceives is orange; rather, it says that Jones perceives that everything he perceptually recognizes is orange. And now we have something which is too weak, for presumably there are plenty of people at the brickworks Jones perceives but fails to recognize (he does not perceive who they all are). So we still haven't been
able to find a way to symbolize what we set out to symbolize without making use of the perceptual quantifier 

Formally speaking, the difficulty is this. A Q-formula involves talking about a system of worldlines each of which weaves its way through the set of compatible worlds. Now since there are individuals in each compatible world which may be untouched by these lines, we must find some way to distinguish between the individuals of a given world that are picked up by a perceptual individual and the individuals which are not. This is essentially what we tried to do by specifying the antecedent 'Ax' in 'Perceives(x)(Ax→x is orange)'. However, there does not seem to be a way to symbolize 'Ax'.

And so we conclude that Q-formulas too are really irreducible to formulas in the language of ordinary quantifiers. When we want to talk about a whole group of perceived individuals at once we really seem to require the resources of Hintikka's new quantifiers. In this way, the same as with P-formulas, the quantifiers do appear to increase the expressibility of Hintikka's system. The value of employing perceptual quantifiers does not seem to lie, as Hintikka believes, in the expression of direct-object constructions. But in order to formally capture those English sentences whose symbolizations are either P-formulas or Q-formulas, the new quantifiers appear to be impossible to eliminate.
1 Hintikka, 'On the Logic of Perception', p. 143.

2 As Hintikka puts it, "...to specify what someone, say a, perceives is to describe what the world is like according to his perceptions. Since these perceptions do not fix the world uniquely, this description is logically speaking not unlike a disjunction of several alternatives concerning the world. The most systematic way of spelling out these several alternatives is to make each of them as full a description of the world as we can give by means of the resources we are using. Switching to an obvious semantical jargon, it is obvious that what such maximal (consistent) descriptions describe is a possible world." Hintikka, 'Information, Causality, and the Logic of Perception', p. 4.

3 A model set w can be said to be compatible with a model set w just in case w is compatible with everything Jones perceives in w.


5 What we say about '(∃x)' and '(∀x)' is not intended to bear upon the similar role these quantifiers play in a semantics for 'Knows' Hintikka has suggested in recent years (see chapter 3).

6 We assume as given rules governing non-modal formulas. For these see Knowledge and Belief (Ithaca: Cornell University Press, 1962).


9 Of course, worlds are not strictly speaking a part of Hintikka's system, but what this comes to in terms of complete novels is obvious.

10 Hintikka, 'Information, Causality, and the Logic of Perception', pp. 4-5.


12 Hintikka, 'On the Logic of Perception', p. 163. Nothing so far guarantees the presence in the model structure of every complete novel which (relative to a given language) can be formed. However, Hintikka seems clearly to assume that this guarantee holds ("to specify what a perceives is to specify the set of all possible worlds compatible with his perceptions", 'Information, Causality, and the Logic of Perception, p. 4). To avoid confusion let us say of a model structure that it's "W-complete" just in case the set W of complete
novels it contains is maximal in this respect. Only now are we justified in asserting that there really is a compatible world in which Tittle is the man in front of Jones.  

15 Ibid., p. 11.  
18 A formula 'virtually implies' another in a Hintikka-type semantics just in case their conditional is 'self-sustaining'. A formula is 'self-sustaining' just in case it belongs to every model set. Two formulas whose biconditional is self-sustaining are 'virtually equivalent'. A formula is 'indefensible' just in case it belongs to no model set. We shall employ this terminology throughout the dissertation.  
19 This accords well with Montague's treatment of 'seems to see'. The sentence "Jones seems to see a unicorn" is translated into the formal language with an operator 'seems' binding the clause 'there exists a unicorn...'. ('On the Nature of Certain Philosophical Entities', p. 179)  
20 Kant, A Critique of Pure Reason, p. 303.  
21 Terry Parsons has suggested there still might be problems with sentences like "Jones perceives that there is something moving" in connection with the recently expressed belief of some scientists that visual perception of movement operates independently of visual perception in general. I have no idea what to say about cases such as these.  
22 This formula too would be inconsistent were the relation R in Hintikka's system to be transitive and reflexive.  
24 We speculated that formulas of the form '(∃x)(Ax & Perceives Bx)' resist the sort of reduction which eliminates '(∃x)', and we may certainly be right. Yet these formulas are incapable of providing a satisfactory formal reading for English sentences, as we have seen, and so it may still well be the case that Hintikka's new quantifiers fail to increase the expressibility of Hintikka's system in any profitable way.  
As Thomason points out, "It is not possible to eliminate the quantifications of a many-individuated logic in favor of predicates and a single quantifier," Thomason, 'Perception and Individuation', p. 279.

Notice that the class of P-formulas and the class of Q-formulas are not mutually exclusive.

There is a minor difficulty here that we shall ignore. There is a difference between (i) Jones's perceiving that what he perceives is F, and (ii) What Jones perceives being such that Jones perceives it to be F. What Jones perceives might well be different from what he perceives that he perceives since R is not transitive and not reflexive. Since the suggestion fails for other reasons we'll pass over this difficulty.
CHAPTER II

A formal semantics for the Logic of Perception bearing considerable similarity to Hintikka's system is proposed by Richmond Thomason in his recent study 'Perception and Individuation'. Like Hintikka's two papers on the Logic of Perception, 'Perception and Individuation' leaves many details of the formal semantics unspecified and makes only the barest mention of such crucial issues as quantifying in and the construction of world lines. Yet the system that does emerge is a fascinating variant of the model set theoretic semantics for perception Hintikka has proposed. In this chapter Thomason's semantics for perception (which we'll refer to as "T") will be presented and discussed, with particular attention paid to the ways in which it differs from Hintikka's system and the desirability or undesirability of these novel features.

* * *

It will perhaps be easiest to approach Thomason's paper by beginning with some remarks of an informal nature and then proceeding to a detailed presentation of the formal semantics. This preliminary discussion can be divided into a series of informal remarks.

1. General structure of T. We have seen that Hintikka's semantics for perception is based upon the notion of a model structure consisting of model sets. Something was perceived to be true just in case it turned out to be a member of every compatible complete novel. With T the situation is different; model sets and complete novels play no part in the system. Rather than defining truth in terms of set membership and rules of deduction in terms of simultaneous
satisfaction, Thomason employs a valuation function which maps well-formed formulas to truth-values. In particular, formulas are mapped to truth-values relative to possible worlds, or "situations" as Thomason more frequently calls them. Formulas are true relative to situations.

Nothing in the way of a radical departure from Hintikka-type semantics seems to hang on this structural feature. Situations bear the compatibility relation to one another and (like Hintikka's complete novels) give us the means to tell what an agent perceives in a given state of affairs. To find out, we simply observe which, of all the situations there could be, are compatible and which are not.

2. The modal operator. The modal operator employed in T constitutes a slight point of departure from Hintikka (although we shall later discover that the point of departure is not as slight as it first appears). Thomason confines his semantics to visual perception and adopts the notion of 'seeing that' as the fundamental perceptual concept in T. An agent 'sees that' such-and-such is the case if and only if such-and-such is the case in every compatible situation (or possible world). The letter "S" is taken to be the modal operator designating the locution 'sees that'.

Thomason nowhere employs the use of subscripts identifying agents. Nor does he introduce more than a single compatibility relation between situations. And hence it looks as though T is a system whose concern is that which a single agent visually perceives. To capture what each of two agents visually perceives two compatibility relations are required to distinguish exactly which states of affairs (since they shall differ) are compatible with what each perceives.
3. **Deductive closure.** A feature shared in common by T and Hintikka's system (perhaps an unfortunate one) is this: As a result of the general strategy employed in the semantics, every logical consequence of what is perceived turns out itself to be perceived. This may easily be shown as follows. Let 'A' and 'B' be well-formed formulas. 'Jones sees that A' is true if and only if 'A' is true in every situation compatible with everything Jones sees. But suppose 'A' logically entails 'B'. Then 'B' is true in every situation logically compatible with everything Jones sees; otherwise in some possible world 'A' is true and 'B' false, which is impossible. But if 'B' is true in every compatible situation, it follows that Jones sees that B.

This phenomenon has been a much-criticized feature of Hintikka-type semantics. But it is admittedly a feature which is not nearly as objectionable in a semantics for perception as in a semantics for knowledge or for belief. To say that a man knows the logical consequences of everything he knows, or believes the logical consequences of everything he believes seems more objectionable than to say a man perceives the logical consequences of what he perceives. Perceiving logical consequences does not seem to involve the extent of mental effort or calculation required to know or believe logical consequences. At any rate, deductive closure is a phenomenon which occurs in T as well as Hintikka's various semantics for propositional attitudes.

4. **Direct-object constructions.** Like Hintikka, English direct-object constructions (like "Schwartz sees the footstool") are handled by Thomason in terms of the propositional 'seeing that'
construction. Also like Hintikka, they are eventually found to be capable of ambiguity. In one sense Jones can be said to see an alligator only if he sees that the object he sees is an alligator; but in a weaker sense Jones sees an alligator when he sees something (that turns out to be an alligator) but does not see that it's an alligator. These seem to correspond (respectively) to Hintikka's *de dicto* and *de re* types of direct-object constructions. However, as we shall see later, the formal treatments of these two constructions in T differ significantly from Hintikka's distinction between 
\((\exists x)(\text{Perceives } b=x)\)' and 
\((\exists x)(b=x \& \text{Perceives}(E y)(y=x))\)'.

5. **Constraints on the compatibility relation.** Perhaps the most striking feature found in T which is absent in Hintikka's Logic of Perception is the introduction of a compatibility relation R which is both reflexive and transitive. As far as transitivity goes, Thomason has no strong feelings. He makes R transitive only "provisionally", claiming that "there are few intuitive resources to mobilize in resolving the question whether this is desirable or not". The reflexivity of R, on the other hand, appears to be a more serious matter.

We're asking whether seeings—that are always successful, so that if SA is true then A is true. As long as we don't confuse 'seeing that' with 'seeming', English usage supports a positive answer to this question; one can't see that a thing is so without it actually being so. If someone claims that he sees that there are pumpkins in the field, he will be forced to withdraw this claim once he has been shown that there aren't pumpkins in the field. In this respect 'sees that' resembles 'knows that', 'realizes that', 'is surprised that', and 'forgets that'.

It will be valid to conclude in T, therefore, that whenever an agent sees that such-and-such is the case, such-and-such is
indeed the case. The actual world is guaranteed to be compatible with itself. And by transitivity it turns out that whenever an agent sees that such-and-such is the case, then he sees that he sees that such-and-such is the case. Together these conditions insure the equivalence in T of 'S(A)' and 'SS(A)', for any well-formed formula 'A'.

6. **Perceptual individuation and perceptual domains.** The most significant feature of all which Thomason's semantics for perception shares in common with Hintikka's is the presence of two independent methods of individuating what is perceived by the agent in question. There are objects which are physically individuated, and there are objects which are perceptually individuated (as with Hintikka, an object b is physically individuated for Jones just in case Jones perceptually recognizes b). And corresponding to these two methods of individuation are two kinds of quantifiers, physical quantifiers and perceptual quantifiers. Notationally, Thomason writes an existential perceptually quantified formula as '(∃θ) Qθ' (where variables are Greek letters) and an existential physically quantified formula as '(∃x) Qx' (where variables are Roman letters). For the sake of simplicity, however, we shall throughout use Hintikka's symbols for quantifiers and thereby hope to avoid confusion in comparing the systems.

The two methods of cross-identifying individuals from world to world nevertheless take on an added complication in T. Each situation (world) has not one, but two domains of individuals (or "individuals_1" as we called them in Chapter One): A domain of physical individuals_1 and a domain of perceptual individuals_1. There are two modes of
existence in T; something can exist physically in a situation w, or it can exist visually. The Buckingham Fountain, for example, exists physically in the actual world, while the Fountain of Youth does not. Yet the latter may very easily become a visual object (i.e., perceptual individual) in the actual world. It often happens, of course, that a given individual may at the same time be both a physical and visual object in the same situation. Yet each domain is independent from the other; some objects may be physical but not visual objects, and some visual objects may fail to be physical objects. As we shall see later, some interesting difficulties stem from this distinction between modes of existence, difficulties which do not arise in Hintikka's semantics for perception.

It might appear at this point as though Thomason's semantics for perception, by virtue of distinguishing physical and visual objects in each situation, will turn out to be suspiciously suggestive of a sense-datum theory. This impression is strongly encouraged by passages like the following:

In a way MacBeth sees something -- he has an impression of a dagger -- and yet there is no physical object with which the dagger can be identified. Here it's plausible to say that MacBeth sees the dagger, so that $'(\exists x)(x=b)'$ will be true, but to deny that the dagger is to be identified with any physical object, so that $'(\exists x)(x=b)'$ will be false.4

Whereas Hintikka's semantics with its perceptual individuals (individuals$_1$) mapping worlds to individuals$_2$ may have looked to a certain extent (as Hintikka himself admits) like a sense-datum theory, there was nothing in the way of "dual domains" in each world. T, on the other hand, looks as though visual sense-data are part of the furniture of each situation, functions or no functions. MacBeth's
dagger just sits in the actual world, as much a member of the visual
domain as MacBeth himself is a member of the physical domain.

Thomason, in section VIII of his paper, readily agrees that his
may be regarded as a sense-datum theory but not in what he calls the
"strong sense". To be a sense-datum theory in the strong sense T
must be able to capture the validity of \( (x)(\forall y)(x \neq y) \). In every
situation in every model structure it must turn out that nothing
qualifies simultaneously as a physical object and visual object.
But, as we've already seen, \( (x)(\forall y)(x \neq y) \) is not valid in T;
hence, if T is a sense-datum theory at all, it is one only in an
innocent sense.

7. The nature of perceptually individuated objects. World lines
are formed in T as they are in Hintikka's system. Thomason describes
them as "rules" which assign to each situation a member of the
universal domain \( 'D' \). \( D \) is defined as the union of all objects,
physical or visual, found in any situation in the model structure
(\( D \) may even -- if it seems desirable -- contain individuals which
do not exist in any situation). The member of \( D \) assigned to some
situation \( a \) by some world line is said to be the "perspective pre-
sented by the \[ world line \], viewed from the situation \( a \].^5 A world
line \( d \) may assign to a situation \( a \) any member of \( D \) whatever, but for
\( d \) to be a physically individuated object the actual situation must
be assigned a member of its physical domain, and for \( d \) to be a
perceptually individuated object the actual situation must be assigned
a member of its visual domain. No such restrictions, however, apply to
situations other than the actual one. We'll examine this interesting
feature shortly.
Physical and perceptual cross-identification over situations comprise what Thomason calls the two "modes of individuation". We shall refer to the modes of individuation as "n" and "m", for the physical and perceptual modes, respectively (Thomason uses the Greek letters "nu" and "mu"). Formally speaking, the modes of individuation are classes whose members are the set of world lines which qualify as individuals. Thus, m is the set of all and only perceptual individuals and n is the set of all and only physical individuals. We then refer to the physical domain of a situation as 'D_a^n', and the visual domain of a as 'D_a^m'. D itself may then be formally defined as ∪_{a ∈ K}(D_a^m ∪ D_a^n), where "K" denotes the set of situations in the model structure.

* * *

We are now ready to offer a formal specification of T. The perceptual model structure consists of a septuple \( \langle K, R, D^m, D^n, m, n, V \rangle \) of elements. We have had occasion already to introduce most of these elements. K is the set of situations, R is a two-place compatibility relation taking situations as arguments, and m and n are the modes of individuation (sets of world lines). We let 'D^m' and 'D^n' name functions which map to each situation its visual and physical domains, respectively. And 'V', finally, names a function which assigns values to individual variables and constants and truth-values to well-formed formulas.

Almost nothing is said of the assignments made by V. It is not clear, for example, how truth-values come to be assigned to atomic formulas. The only truth-value assignments specified are the assignments to quantified formulas. Once again using Hintikka's...
notation, these quantifier rules for $T$ come to the following:

$$V_a((\forall x)(Ax)) = T \text{ if and only if } (\forall^d x)_a(A) = T$$  
for all $d$ belonging to $m$, such that $d(a)$ belongs to $D^m(a)$

$$V_a((x)(Ax)) = T \text{ if and only if } (\forall^d x)_a(A) = T$$  
for all $d$ belonging to $n$, such that $d(a)$ belongs to $D^n(a)$.  

Here we let the expression "$\forall^d x$" denote the function which is exactly like $V$ except for the assignment of $d$ to $x$. And "$d(a)$" stands for the member of $U$ assigned a situation $a$ by the world line $d$.

What these rules say intuitively is this. A perceptual universally quantified formula '$(\forall x)(Ax)$' is true in a situation $a$ just in case 'Ax' is satisfied by every world line which belongs to the set $m$ and which assigns to $a$ a member of its visual domain. A physically universally quantified formula '(x)(Ax)' is true in a situation $a$ just in case 'Ax' is satisfied by every world line which belongs to $n$ and which assigns to $a$ a member of its physical domain. From these rules we may immediately derive rules for the existential quantifiers as follows:

$$V_a((\exists x)(Ax)) = T \text{ if and only if } (\forall^d x)_a(A) = T$$  
for some $d$ belonging to $n$, such that $d(a)$ belong to $D^n(a)$

$$V_a((\exists x)(Ax)) = T \text{ if and only if } (\forall^d x)_a(A) = T$$  
for some $d$ belonging to $m$, such that $d(a)$ belongs to $D^m(a)$.

There is much that is interesting about these rules. In the first place, the truth of a formula like '(Ex)S(x=Smith)' is dependent not only upon the world line picking Smith out of each situation but upon the presence of Smith in the actual world. In Hintikka's semantics, we may recall, no such restriction was present. One could perceive who an individual is even though no such individual of
happened to exist in the actual world. Hence the rules for ordinary quantification in T are in an important way different from Hintikka's. This difference, it should be noted, stands independent of reflexivity. Hintikka's system could lack the restriction even with the reflexivity feature added.

Interesting too is the absence (which we alluded to earlier) of a requirement that the world line pick from each situation (over which it's defined) a member from a specified domain in that situation. For example, the formula '(Ex)S(x=Smith)' is true in a if and only if there is a world line d which (i) Satisfies 'S(x=Smith)', (ii) Is a member of n, and (iii) Assigns a a member of it's physical domain. But there's no requirement to the effect that d pick out from each situation a member of its physical domain. Only in a must Smith be a physically existing object. Similarly, the world line which satisfies 'S(x=Smith)' when Smith is not perceptually recognized (namely, a member of m) must assign a a member of its visual domain but may conceivably assign other situations member of their physical domains.

Moving now to a more detailed consideration of the modes of individuation m and n, we find two rather fascinating requirements concerning their membership (see 5.2 on page 272). Since Thomason decides in the end to reject the second, we'll concentrate our attention upon the first, viz., no two members of the same mode of individuation may assign to the same situation the same element of D. That is, if a is a situation, and if d₁ and d₂ are both elements of either m or n, then if d₁#d₂, d₁(a)#d₂(a). "These objects," Thomason writes in reference to world lines in either m or n, "must be chosen systematically, so that they do not intersect: two different objects of the same mode
individuation cannot present the same perspective in the same situation."®

This restriction is rather interesting and on the surface appears to be quite reasonable. Suppose that Jones sees that two men are standing in front of him but sees who neither man is. Suppose in addition that it's compatible with everything he sees that either is his father. Then in at least one compatible situation the one man is his father, and in at least one compatible situation the other is his father. But according to the restriction there is no situation in which both are his father. In no situation are the two men identical, and this result seems to square well with our intuitions.

As a matter of fact, the restriction Thomason introduces here is a very strong one. Not only does it guarantee the validity of 

\[ (x) (y) (x = y \rightarrow S(x = y)) \]

and

\[ (\forall x) (\forall y) (x = y \rightarrow S(x = y)) \]

in T, but much more surprising, the validity of 

\[ (x) (y) (x \neq y \rightarrow S(x \neq y)) \]

and

\[ (\forall x) (\forall y) (x \neq y \rightarrow S(x \neq y)) \].

We shall demonstrate informally the validity of each. First, suppose 

\[ \neg (x) (y) (x = y \rightarrow S(x = y)) \]

is true in some situation \( a \) (relative to some model structure \( \langle K, R, D^n, D^m, m, n, \mathcal{V} \rangle \)). Then

\[ (\exists x) (\exists y) (x = y \& -S(x = y)) \]

is true in \( a \). So \( V_a ( (\exists x) (\exists y) (x = y \& -S(x = y)) ) = T \). And by two applications of the quantifier rules,

\[ ((\forall x) (\forall y) (x = y \& -S(x = y)) ) = T, \]

for some \( d_1 \) and \( d_2 \) in \( n \), such that \( d_1(a) \) and \( d_2(a) \) belong to \( D^n(a) \). But from this it follows that both

\[ ((\forall x) (\forall y) (x = y )) = T \]

and

\[ ((\forall x) (\forall y) (x \neq y )) = T, \]

for some \( b \) such that \( R_{ab} \). But this violates the restriction that distinct world lines may not intersect anywhere (as they do here in \( a \)). Thus, 

\[ (x) (y) (x = y \rightarrow S(x = y)) \]

is valid. The proof of the validity of

\[ (\forall x) (\forall y) (x = y \rightarrow S(x = y)) \]

is nearly identical.

Now suppose 

\[ (x) (y) (x \neq y \rightarrow S(x \neq y)) \]

is false in some situation \( a \)
Then 

\[ V_a((Ex)(Ey)(x\neq y \& \neg S(x\neq y))) = T. \]

So, like the above, 

\[ ((V^{d_1/x}d_2/y)_a(x\neq y \& \neg S(x\neq y))) = T, \]

for some \( d_1 \) and \( d_2 \) in \( n \), such that \( d_1(a) \) and \( d_2(a) \) belong to \( D^n(a) \). But then 

\[ ((V^{d_1/x}d_2/y)_a(x\neq y) = T \] and \( ((V^{d_1/x}d_2/y)_b(x=y) = T, \]

for some \( b \) such that \( Rb \). And again Thomason's restriction is violated. Two distinct world lines merge when they reach \( b \), and so we conclude that 

\[ '((x)(y)(x\neq y \ S(x\neq y)))' \]

is valid in \( T \) with the restriction added. Likewise, 

\[ '((\forall x)(\forall y)(x\neq y \rightarrow S(x\neq y)))' \]

is valid.

The formulas 

\[ '((x)(y)(x=y \rightarrow S(x=y)))' \] and 

\[ '((\forall x)(\forall y)(x=y \rightarrow S(x=y)))' \]

guarantee that no world line may split when it extends from a world to its alternatives. And the formulas 

\[ '((x)(y)(x\neq y \rightarrow S(x\neq y)))' \] and 

\[ '((\forall x)(\forall y)(x\neq y \rightarrow S(x\neq y)))' \]

guarantee that no distinct pair of world lines may merge when they extend from a world line to its alternatives.

It is interesting to see what Hintikka has to say about this particular sort of phenomenon.

The structure formed by the relations of cross-world identity may be so complex as to be indescribable by speaking simply of partial identities between the domains of individuals of the different possible worlds. Above, it was said that in the case of many propositional attitudes an individual cannot 'split' when we move from a world to its alternatives. Although this seems to me to be the case with all propositional attitudes I have studied in any detail, it is not quite clear to me precisely why this should always be the case. At any rate, there seems to be reasons for suspecting that the opposite 'irregularity' can occasionally take place with some modalities: individuals can 'merge' together when we move from a world to its alternatives.9

Starting with the no-splitting condition, we find that in Hintikka's (revised) semantics for knowledge the formula 

\[ '(x)(y)(x=y \rightarrow \text{Knows}(x=y))' \]

is valid (self-sustaining). But the no-merging condition is provable in none of Hintikka's systems. A Hintikka-
type semantics having a symmetric compatibility relation as well as a rule like

\[
\begin{align*}
(Ex)\text{Knows}(b=x) & \subseteq w \\
(Ex)\text{Knows}(c=x) & \subseteq w \\
\frac{b=c}{\text{Knows}(b=c)} & \subseteq w
\end{align*}
\]

would suffice to capture the no-merging condition, but none of his systems have symmetry. In a system with symmetry it turns out that anything compatible with what is \( \phi \)'d, where \( \phi \) is the propositional attitude in question, is true; and to avoid this result symmetry is never considered.

Merging is a phenomenon which, Hintikka believes, can occasionally take place in semantics for propositional attitudes. By this Hintikka appears to be saying that in actual situations (real-life situations) a no-merging condition sometimes appears to be violated; in knowledge, for example, it seems dubious to conclude that an agent always knows that two individuals \( x \) and \( y \) are distinct just in case (i) He knows who \( x \) is, (ii) He knows who \( y \) is, and (iii) \( x \neq y \). And if the formal semantics for knowledge is to reflect ordinary knowledge, it's not clear that world lines should not be allowed to merge.

What about perception? Is Thomason's restriction going to be at odds with ordinary perceptual circumstances? Can a man see who \( x \) is, see who \( y \) is, and fail to see that \( x \neq y \) when \( x \) and \( y \) are in fact distinct? At first it seems as though this is impossible. If \( x \) and \( y \) are both before Jones's eyes, and if Jones sees who each is, and if \( x \neq y \), it seems that Jones cannot possibly fail to see that they're distinct. But the matter is not quite so simple. As we shall see
later, it is not the case in Thomason's semantics that if Jones sees who x is, Jones sees x. To see who a man is it is not necessary for him to be present to your immediate senses, according to Thomason, and from this fact it does begin to look as though the no-merging restriction may be overly restrictive after all from the point of view of ordinary perception. We shall return to this matter in the next section.

* * *

Judging from what both Hintikka and Thomason have to say about physically individuated objects (namely, that to be physically individuated an object must be perceptually recognized) it would not seem unreasonable to conclude that in both of their systems the set of physically individuated objects will turn out to be a proper subset of the set of perceptually individuated objects. The reason is simple. If a man perceives who Smith is, then surely the man must be perceiving Smith. A man cannot perceptually recognize Smith, it would seem, if he fails to perceive Smith in the first place. Thus, it would appear that to qualify as a physically individuated object an object must already be a perceptually individuated object.

Interestingly, it is just this sort of reasoning Thomason wishes to attack.

Suppose we're at a cocktail party and there is someone across the room you're trying to point out to me... Suppose that from your description I realize who he must be; he's a person well known to me whom I know to be at the party. But the light is bad, people are milling around, and I haven't been able to pick him out from the crowd. The natural thing to say in these circumstances is
(3.17) I see who you mean but I don't see him

... The formalization of (3.17) is

(3.19) (Ex)S(x=a) & -((3 x)S(x=a)

\[ \text{[where 'a' stands for 'the person you mean']}. \]

In this example the agent perceives who \( a \) is without perceiving \( a \), and this looks like a counter-example to the claim that perceiving who implies perceiving.

At first glance this example does not seem convincing at all. It seems rather clear in the example that I know who \( a \) is. There is someone of whom I know that he's the person you mean. But do I visually perceive who he is? Is there really someone of whom I perceive that he's the person you mean? It's hard to see that there is. On the face of things there seems to be a confusion here between my knowing who \( a \) is and my visually perceiving who \( a \) is in this example. It's not clear that because I know who \( a \) is I thereby see who \( a \) is.

It nonetheless becomes clear later in the paper that Thomason is not confusing 'knowing who' with 'seeing who' at all. The whole matter is actually much more subtle than what it appears to be.

(6.1) (3 x)(x=a)

(6.4) (Ex)S(x=a),

and suppose that we are concerned with what is seen by an agent \( p \).

Notice that 6.4 can be true when \( p \) is not looking at the object referred to by 'a': that is, 6.4 does not imply 6.1. All that is needed for 6.4 to be true is that the object should exist and be physically the same in all situations compatible with what \( p \) sees. Since in English we commonly use locutions such as 'seeing what' or 'seeing who' to express 6.4, this fact helps to explain how 'seeing' can take on non-visual overtones akin to 'knowing'. For me to see who the man is, it is not necessary for me to be looking at him; the realization may strike me while my back is turned to him. Thus, 6.4 does not imply 6.1.
To this Thomason adds by way of footnote:

For some reason, many philosophers with whom I have discussed this claim have found it peculiar and somehow objectionable. Nevertheless, it's the way we all use the language: 'seeing that' and 'seeing who' are conditioned by knowledge, without regard to how this knowledge was obtained. Someone who is blind can "see that" and "see who" as well as anyone else, and no metaphors are involved here.  

It comes as rather a shock to advance to this point in 'Perception and Individuation' and suddenly learn that the 'sees that' locution we have been discussing all along is to be understood in such a way that a blind man can 'see that' such-and-such is the case. Small wonder then that I can see who someone is without actually seeing him at a given moment. A whole semantics for knowledge has suddenly eased its way in the back door. As far as the cocktail party is concerned, I see who you mean just insofar as I know who you mean. If I know who someone is, then I automatically see who he is in some sense of the word "see". And if I know something is true, then I automatically see that it is true in that same sense of the word "see". Apparently it's just this sense with which Thomason has been concerned all along, and we were mistaken to have ever thought otherwise.

There is, of course, nothing objectionable about offering a semantics for a sense of "sees that" which is such that knowing that entails seeing that. It may even prove advantageous in reflecting more accurately the actual use of English expressions (an area in which Hintikka is often criticized). But whatever is gained in paralleling everyday discourse is lost, I believe, in the failure to present a clear and unified account of a single, unambiguous concept.
For T now looks as though it combines into a single system a semantics for visual perception and a semantics for knowledge, and it does so with the use of a single modal operator. Precisely the same system is supposed to elucidate the use of "sees" in "Smith sees a woodpecker" and "Smith sees now that he'll soon lose his job".

Suppose we say that an agent "visually perceives" that something is the case if and only if his eyes actually report to him that it's the case. Then "visually perceives that" is the meaning of "sees that" which we originally took to be the interpretation of the modal operator 'S'. What we seem to have learned since is that in reality an agent "sees that" something is the case if and only if either he visually perceives that it is or else knows that it is. 'Sp' is true if and only if either 'p' is compatible with everything the agent visually perceives or with everything the agent knows.

This interpretation of "seeing that" brings with it an important disadvantage. Every formula having modal operators will be systematically ambiguous between what's known and what's immediately perceived; and, as a result, there appears to be no way in which immediate perceptions can be captured in the formal semantics. If Jones visually perceives that a man is standing in front of him, it won't do to suppose 'S(Ex)(Mx & Fx)' is the correct symbolization. This is too weak. It says that Jones sees that a man is standing in front of him, and this leaves open the possibility that Jones might only know it. The formula might better be read, "Jones either knows or visually perceives that a man is standing in front of him". There just seems to be no formula in T capable of expressing "Jones visually perceives that a man is standing in front of him".
Having seen that Thomason has chosen to interpret "sees that" in this wider fashion, we return to Hintikka's suspicions in 'Semantics for Propositional Attitudes' that it may be unrealistic to impose a no-merging restriction on world lines of a semantics for certain propositional attitudes. Earlier we noted that a no-merging restriction is not clearly desirable to impose upon a system of knowledge. It now begins to look as though "sees that" is no different. The reason is very simple.

If Jones knows who each of two people b and c are, then '(Ex)(Ey)S(b=x & c=y)' is true even when neither b nor c call within the field of Jones's vision. But if b and c are really distinct it follows by the no-merging restriction that 'S(b≠c)' is true in T. But 'S(b≠c)' is true just in case Jones either visually perceives that b≠c or knows that b≠c. Since the former is impossible when b and c are not within his visual field, it follows that if 'S(b≠c)' is to be true in T, Jones must know that b≠c. But concluding from all of this that Jones really does know that b≠c seems rash, as we've already indicated. And so if there is uncertainty (as there surely seems to be) in imposing the no-merging condition upon a semantics for knowledge, precisely the same uncertainty extends to T.

But one ought to go on to ask whether the no-merging condition '(x)(y)(x=y → S(x=y))' is any more unrealistic in this regard than the no-splitting condition. Is there anything more unrealistic in insisting that Jones know that distinct individuals known to him are distinct than to insist that identical individuals known to him are identical? It's hard to see any difference here. It has in effect been suggested by Casteneda (See Chapter Three) that the no-splitting
condition is violated in ordinary epistemic circumstances. An amnesiac may be the famed War Hero wounded 100 times in a recent war, he may know who the War Hero is (through research), he knows who he himself is, and still he may fail to know that he is in fact the War Hero (no one knows what became of the hero). Casteneda's may not be as convincing as hosts of other examples one might devise, but it seems to show that '(x)(y)(x=y→Knows(x=y))' is at least as questionable a validity to have in one's system as '(x)(y)(x≠y→Knows(x≠y))'. Neither seems desirable.

Then why does Hintikka not fuss over splitting? No doubt because it cannot be avoided in his system (as we shall demonstrate in the next chapter). There is no corresponding compelling reason to make '(x)(y)(x≠y→Knows(x≠y))' a validity, however, and so Hintikka hesitates incorporating anything along the lines of a no-merging restriction. All things considered, it seems to me that Hintikka is not much better off than Thomason. Both employ restrictions on world lines that do not clearly reflect the way people actually know or perceive things to be.

* * *

It was mentioned earlier that a major departure from Hintikka's semantics for perception was the presence of reflexivity in T. If an agent sees that something is the case, then it follows that it is indeed the case. Every situation in the model structure has itself as an automatic alternative. In Hintikka's semantics, on the other hand, it is often the case that a world fails to be compatible with itself. An agent can perceive that something is the case without it's being the case.
We recall from our discussion of Hintikka, however, that reflexivity could just as well have been employed in his system. It wasn't employed for the reason that it would have, according to Hintikka, produced a system which would be substantially less interesting. It would not have allowed reference to illusions, hallucinations, perceptual mistakes, and the like. If someone can be said to 'perceive' something to be true only if it's really true, how can we talk of perceptual mistakes without increasing the resources of the language? It would seem that reflexivity precludes expression of Jones's illusions, hallucinations, and perceptual mistakes.

Won't it be, then, that Thomason's system T will have troubles of a similar nature capturing situations in which a person incorrectly views his actual environment? If someone can be said to "see that" something is the case only if it really is the case, it seems as though illusions and incorrect perceptions cannot be handled by T.

In view of this, it is curious that Thomason discusses examples which involve just these sorts of perceptual mistakes. He considers, for example, a case in which a man looks at a cluster of leaves and has an impression of a deer. And in another example (which we've already mentioned) MacBeth hallucinates and has the impression of a dagger when no dagger is present (and where in addition there is no physical object for which the so-called dagger could be mistaken). What he says about these examples reveals that to a certain extent perceptual mistakes can be captured in the system after all. They cannot be captured in terms of the "seeing that" idiom, it is true, but that will make no difference. For as we may well have forgotten
by this time, we do have at our disposal a domain \( D^n(a) \) of visual objects in each situation \( a \). And we finally have occasion to use these resources.

The MacBeth case is quite straightforward. Although \( 'S(Ex) \) (\( x \) is a dagger)' is false in the case under consideration (and would be true in Hintikka's system), \( '(\exists x)(x \text{ is a dagger})' \) is not. There is a dagger found in the domain of the actual world. Hence \( '(\exists x)S(x \text{ is a dagger})' \) is true, and we have \( '(\exists x)(S(x \text{ is a dagger}) \& -(\forall y)(x=y))' \) as the correct symbolization of the Macbeth example. And in the example where the cluster of leaves resembles a deer, we have a situation in which a visual object, the deer, is identical to a physical object, the cluster (see page 274 of Thomason's paper). And so \( '(\exists x)(S(x \text{ is a deer}) \& (\forall y)(y \text{ is a cluster} \& x=y))' \) is the symbolization we are after.

There finally seems to be a worthwhile purpose served by the presence in each situation of dual domains. What appeared a perhaps unnecessary complication in the system turns out in the end to cleverly provide a means of capturing hallucinations. A hallucination does not emerge in the system as a world line (as for Hintikka) but as an element in the visual domain of a single situation. The visual domain will then consist of three kinds of objects: Those which satisfy an open formula of the form \( 'S(Ex)(x=y)' \), those (like the deer) which satisfy \( '(Ex)(x=y)' \) but nothing of the form \( 'S(Ex)(x=y)' \), and those (hallucinations) which satisfy neither. The latter two kinds of visual objects seem to provide the handle we need to talk about perceptual mistakes in T.

Will this suffice to give formal expression to every kind of
perceptual mistake an individual might commit? I am not convinced that it does. Suppose that Smith is standing in front of Jones and that Smith appears to Jones to have black teeth. But suppose that in actuality Smith has very white teeth and only happens to be chewing on licorice. Can the sentence, "Smith seems to Jones to have black teeth" be captured in T? Will the presence of visual objects enable us to talk about Smith's mistaken perception?

The natural move might be to suppose that our visual domain includes a character we might refer to as "Black Tooth Smith", an individual just like Smith except for the color of his teeth. We can then perhaps symbolize the sentence "Smith seems to Jones to have black teeth" as \((\exists x)(x=\text{Black Tooth Smith} \& S(\exists x) \& (\forall y)(y=\text{Smith} \& x=y))\). But this surely won't do if the Indiscernibility of Identicals is to be preserved in T, for Smith cannot be identical to something whose teeth are of different color from his own. There cannot be a member common to two domains in a world, it seems clear, if it is supposed to have different properties in each domain.

For much the same reason Thomason's deer example seems unsettling. Is the physical object in question -- the cluster of leaves -- supposed to be identical to a visual object, a deer? It's difficult to see in what sense the same object can be at the same time a cluster of leaves and a deer. Perhaps, then, the cluster is to be identified with an apparent deer. But what do we mean by "apparent" here? If we were able to supply a formal account of "apparent" in terms of the resources of T, the problem of capturing cases of mistaken perceptions would never have been a worry in the first place.
Perhaps the visual object in question is not the deer at all; perhaps the cluster is both the physical and visual object. There seem to be grounds for this interpretation on the basis of Thomason's claim on page 274 that if 'a' names the cluster of leaves 'a' names a visual object. Then maybe to say that Jones sees that this object is moving, all we need is '(∃x)(Ey)(x is a cluster & x=y & S(x is moving))'. But this is obviously too weak; it tells us nothing about the agent's mistaking the cluster for a deer. It could just as easily be true in a situation where the agent mistakes the cluster for a mosquito. How can we express the deer's role in this example as the apparent object?

We can of course reverse tactics and admit the diversity of Smith and Black Tooth Smith (as well as the diversity of the cluster and the deer). We can imagine that while Smith is a member of D^n(a), Black Tooth Smith is a member of D^m(a), and while Smith and Black Tooth Smith share an almost uncanny resemblance, they are entirely distinct (we can ever speculate that there may be a casual relationship which holds between the two, but that sort of speculation goes beyond the scope of anything the theory is able to handle). Unfortunately, this solution seems to reduce Black Tooth Smith to the position of a hallucination, much like MacBeth's dagger. He can no longer be identified with some physical object. We lose the distinction between mistaking a real thing for something it is not and seeing something where there is nothing.

It would be nice to say in the Smith example we mistake something for something else, that Black Tooth Smith can be identified with a member of the physical domain D^n(a). But what exactly do we mean by
"identified with"? As we have seen, the identity sign cannot be used to identify physical with visual objects (in cases like these). But it's hard to find any clear alternative sense of "identify with" expressible in T that will solve these puzzles. And so we tentatively conclude that T is a system which, unlike Hintikka's, is incapable of accounting for certain types of mistaken perceptions. Hallucinations work out beautifully in T, but nothing else seems to. Cases in which a man mistakes one thing for another appears, on Thomason's account, to create significant problems in the theory. There seems to be no way to indicate in formal terms in what relation Smith stands to Black Tooth Smith or the cluster stands to the deer.

Difficulties of this sort are not uncommon in philosophical theories of perception. It is convenient to postulate sense-data in an effort to explain the similarity involved in perceiving a real thing and perceiving a hallucination. But well-known difficulties arise when one tries to explain precisely the relation that obtains between sense-data and real things. It is not at all clear what relation obtains between a real bear and my sensation of the bear if we construe the latter as a sense datum.

None of this is intended to suggest that there are inherent problems in the postulation of separate domains in each situation. Nor is there any obvious problem in regarding these domains as intersecting; if Jones sees Smith as he really is, then the physical Smith and the visual Smith are supposed to be the same, and that does not seem unreasonable. As we have already seen, the identity between members of $D^M(a)$ and $D^N(a)$ shows that T, if it's a sense-datum
theory at all, is not a sense-datum theory in what Thomason calls
the "strong sense". Visual objects need not be two-dimensional
discs or wave-like images but may be (and usually are) physical
objects.

Though there are no obvious inherent problems with dual
domains, it does begin to look as though they are unnecessary com-
plications in the theory after all. If the presence of visual domains
was able to account for mistaken perceptions in a system whose
compatibility relation is reflexive, then they would serve a useful
purpose. If, however, they account for some cases of mistaken
perceptions and not others, then the virtue of their presence is
questionable. A semantics for perception seems cumbersome enough
with perceptual world lines (members of m) and quantifiers which
range over them. Adding to this special visual domains seems to
create an unnecessary burden in the semantics.

Thomason does seem to suggest, however, another function which
visual objects are supposed to perform -- to give expression to what
we might call "accidental perceptions". Suppose Socrates sees the
man in front of him who is his father, but his father is wearing a mask,
and hence Socrates does not see that the man he sees is his father.

When Socrates denies that he sees his father there is
a sense in which what he says is true. Namely,
(6.9) - (∃x)S(b=x)
is true, where 'b' stands for 'Socrates's Father'....
On the other hand there is also a sense in which what
the sophist says is true when he claims that Socrates
does see his father:
(6.10) (∃x)(b=x)
is true.10

Since we don't wish to say that Socrates both sees and does not
see his father, we turn to the formal semantics and find that we
have the expressive resources to show that there is no contradiction. And the sense in which Socrates does see his father is captured by showing that his father can be found in the visual domain of the actual world. Thus, visual domains do perhaps play a necessary role in the system. They enable one to say that he sees an F but does not see that it is an F.

Nevertheless, even here visual domains seem dispensable. This weaker sense of seeing something is perfectly capable of being expressed without appeal to visual domains. We need only look as far as Hintikka's distinction between \((\exists x)\text{(Perceives } b=x)\)' and \((\exists x)\text{(b=x \& Perceives(Ey)(x=y))}' to convince ourselves that perceptual world lines are adequate to capture the distinction between the two senses in which Socrates might be said to see his father. In the weaker of these senses there is a world line which, whatever else it does, picks from the actual world (out of the one and only domain it has) Socrates's father. And this allows us to say that Socrates sees someone who turns out to be his father: \((\exists x)\text{(b=x \& S((Ey)(x=y)))}'.

In this respect T is even better off than Hintikka's system; that is, perceptually quantified formulas which bind variables occurring both within and without modal operators are much easier to account for semantically in T. The problem we face in Hintikka's semantics, we recall, involves the failure of the actual world to be its own perceptual alternative, and consequently in many cases a puzzle arises as to which element in the actual world we attach the chain. Hintikka's solution (find out what caused our perception) we found impossible to describe in the language of the formal semantics. But no such difficulty arises at all in T. If Socrates
sees his father, his father is guaranteed to show up in the actual world. And in general, whenever we have to evaluate the truth-conditions for a formula of the form \((\exists x)(Ax \& S(Ex))\), the procedure is straightforward. Formulas like \((\exists x)(x \text{ is fat} \& S(x \text{ is not fat}))\) are categorically false and no fuss need be made over them.

* * *

In this chapter I have discussed Thomason's semantics for the locution "sees that". Though his system bears considerable similarity to Hintikka's semantics for "perceives", there are considerable differences. I have argued in some cases these differences represent improvements over Hintikka's system, but there are serious difficulties in other respects. There seems to be no way to symbolize such sentences as "Smith seems to Jones to have black teeth"; there seems to be no way to indicate formally the distinction between visually perceiving that something is true and knowing that it is; there are visual domains in each possible world which seem to serve no useful purpose in the system; and there are constraints on the behaviour of world lines (that they may not intersect) which seem much too restrictive.
FOOTNOTES

1Thomason, 'Perception and Individuation', p. 263.

2Tbid., pp. 263-264.

3For suppose 'S(A)' is true in a situation w. Then in every situation w compatible with w, 'A' is true. But by transitivity 'A' is also true in every situation w" compatible with any situation compatible with w. But 'SS(A)' is true in w if and only if 'A' is true in every such w. Hence 'SS(A)' is true in w.


5Tbid., p. 272.

6Nevertheless, the system Q3 presented by Thomason in his paper 'Modal Logic and Metaphysics' provides some clues along these lines.

7There the restriction was present, of course, in formulas where variables occurring free within and without the scope of the modal operator were bound by '(Ex)'.


10There is no indication that a similar condition is supposed to show up in his semantics for perception; but I believe the same reasons which necessitated his adoption of the rule (C.Ind. =), without which the proof of '[(x)(y)(x=y→Knows(x=y))]' is impossible, likewise necessitate the adoption of a rule which in effect would guarantee a no-splitting condition for physically individuated world lines. For the reasons which necessitated adopting (C.Inc. =), see the next chapter.


12Tbid., p. 275.

13Tbid., p. 275, n. 19.

14Thomason writes, "In interpreting S the modal operator we are considering situations compatible with what is seen, not just with what is known" (p. 275). Compatibility with either an agent knows or everything he visually perceives thus determines the situations related by R in T. Hence knowing p entails seeing that p in the sense of "sees that" Thomason is after.
It can be avoided if Hintikka's system lacked the so-called "Restricted Range feature" (which we discuss in Chapter Four). Dropping this feature seems to be what Castaneda intended to suggest in offering his War Hero example.

CHAPTER III

We have already alluded to Hintikka's move in recent years to incorporate into his model-theoretic semantics for knowledge a second set of quantifiers '$(\neg x)' and '$(\lor x)$. In this chapter we shall (i) Examine in some detail Hintikka's motivation for this change, (ii) Discuss the nature of the expanded system (which we'll call KBC), and (iii) Point out a number of problems which beset KBC.

***

The matter of introducing additional quantifiers to Hintikka's original semantics for epistemic logic (developed in Knowledge and Belief) has a rather unusual history. A now-famous and somewhat bizarre imaginary example offered by Castaneda figures prominently in this history, and in the end it is precisely this bizarre example which drives Hintikka to the expanded system. There is thus no better way, I believe, to examine Hintikka's motivations for introducing new quantifiers than by starting from the beginning and tracing the history of the problem; doing this will prove valuable too in understanding the roles of the new quantifiers in KBC.

I propose to begin by going all the way back to Knowledge and Belief. It is a well-known feature of Hintikka's original model set theoretic semantics for knowledge (KB) that '(Ex)Ka(b=x)' is supposed to be a formal reading for "a knows who (or what) b is" or "Someone (or something) is known by a to be identical to b". Now the English locution "knows who" is notoriously vague and seems to have a variety of uses in English. One is therefore inclined to
feel suspicious from the outset when it is proposed that "a knows who b is" be unambiguously rendered by the single formula '(Ex)Ka (b=x)' in the formal semantics. One wonders how a single formula is supposed to capture at once all the ways in which "knows who" is employed in English.

Hintikka himself seems to anticipate this sort of reaction arising among his readers. Although he writes on page 131 (of Knowledge and Belief) that "knows who" is "easy to translate into our symbolism", he admits later that

The criteria as to when one may be said to know who this or that man is are highly variable. Sometimes knowing the name of the person in question suffices; sometimes it does not. Often "acquaintance" of some sort is required.\(^1\)

Here he agrees that necessary and sufficient conditions for truly saying of a person that a knows who he is do not remain constant from situation to situation. We cannot offer an analysis of the locution "knows who" which applies with complete generality to all our uses of the expression. Yet Hintikka goes on in the same passage to say, "Our discussion is independent of this difficulty, however". How he came to this conclusion at the time I fail to understand, but his addition of '(∃x)' and '(∀x)' to the system some years later is testimony (for reasons we shall explain) that his position on the matter has changed drastically. The very difficulty he speaks of in the passage above turns out in the end to necessitate significant changes in his semantics for knowledge.

Perhaps Hintikka was unhappy with his univocal treatment of "knows who" as early as 1962 (when Knowledge and Belief first appeared). But it should be pointed out here that surely there was never hope that Hintikka's system would provide an exact parallel
with the way in which English speakers use the word "knows". Such was never his project. It was clear at the outset, for example, that deductive closure characterized KB. Thus the English "Jones knows p but not (((p v (q & p)) & (P v q)) v (r & (p v q)))" has no consistent formulation in KB even though the corresponding English sentence is perfectly consistent.

In Castaneda's 1964 review of Knowledge and Belief in the Journal of Symbolic Logic the point is made that formulas of the form '(Ex)Ka(a=x)' (a's knowing who a is) are particularly troublesome in terms of symbolizing the English locution "knows who".

In Hintikka's epistemic system there just seems to be no way of consistently formulating, e.g., the contingent statement

(A) There is a person x such that a knows that a=x but does not know that he himself = x.

In the reviewer's opinion the way of dealing with the logic of self-knowledge is by means of a special descriptor operator "d" such that if a is an individual sign with no occurrences of "d", then "da" is an individual sign. The desired interpretation is shown by symbolizing (A) as '(Ex)(Ka(x=a) & -Ka(x=da))'.

Castaneda argues here that '(Ex)Ka(a=x)' looks as though it is supposed to symbolize two non-equivalent English sentences, "a knows who a is" and "a knows that he himself is a". But, if so, then there is no conceivable way of rendering sentence (A) consistently in the formal language. Hence Hintikka's system is inadequate as it stands.

Curiously, Castaneda does not conclude from this inadequacy in Hintikka's system that there is anything inherently problematic about Hintikka's treatment of "knows who". He suggests remedying the problem by introducing to the language of Hintikka's system an (apparently primitive) operator "d" to formally distinguish between the two sorts of self-reference at issue. Castaneda himself finds this particular remedy ineffective in a later paper, but what's important to note here is his acceptance of the open sentence schema '(Ex)Ka(x=y)' as properly capturing all uses of "knows who". For Castaneda it was just a question of distinguishing
between '(Ex)Ka(a=x)' and '(Ex)Ka(σ a=x)'.

At any rate, the apparent failure of KB to provide a consistent symbolization of (A) without recourse to enriching the language seemed to be a significant criticism of the system. Hintikka's 1966 paper, "Knowing Oneself" and Other Problems in Epistemic Logic, is a reply to this and other criticisms of KB raised by Castaneda in his review. As far as the present difficulty is concerned Hintikka had this to say:

Intuitively speaking, it might seem that an implication from "a's knowing something about a while knowing who a is" to "a's knowing something about himself" is not unproblematic in that it is based upon an implicit assumption. If the assumption is made explicit, however, it is easy to see that it is in fact satisfied. What one has to assume is, loosely speaking, that if someone knows who a is, and if he is in fact himself a, then he knows that he himself is a. This it seems to me is surely the case.... That there is some confusion in Castaneda is shown by his comment that in my system "there just seems to be no way of consistently formulating, e.g., the contingent statement (A)...." On the interpretation I have argued for, however, (A) is clearly inconsistent (indefensible). Since the bound variable x occurs in (A) within the scope of the epistemic operator "Ka", it has to range over such individuals only as are known to the referent of 'a' -- known not merely in the sense that he knows that they exist but also in the stronger sense that their identity is known to him. But if the person in question knows that a refers to one definite individual of this sort, and if this individual is in fact himself, how can he possibly fail to know that 'a' in fact refers to himself? Thus (a) is intuitively inconsistent, and this fact is reflected by its being inconsistent on the assumptions which are made in Knowledge and Belief."

Here Hintikka replies to Castaneda's objection by (i) Admitting that (A) can be symbolized in KB only by an indefensible formula, and (ii) Arguing that (A) is intuitively inconsistent and hence ought to be symbolized only by means of an indefensible formula. Thus "a knows who a is" and "a knows that he himself is a" are both
symbolized '[(Ex)Ka(a=x)]', and this is as it ought to be. Neither sentence can be true while the other is false; they are intuitively equivalent.

Castaneda replied to this in his 1967 paper, 'On the Logic of Self-Knowledge'. Taking issue to Hintikka's claim that (A) is intuitively inconsistent, he sets out here to produce a possible state of affairs in which an agent a knows who a is and also knows who he himself is but fails to know that he himself is a. And so we have the famous War Hero example:

Suppose that X is brought unconscious to a certain military tent; that on gaining consciousness X suffers from amnesia and during the next months becomes a war hero and gets lost in combat and forgets the military chapter of his life, and that later X studies all the accounts of the war hero and discovers that the hero was, e.g., the only one in the war wounded 100 times. For many normal situations, e.g., passing a history examination, X knows who the hero was. In many such situations "There exists a man known to the war hero wounded 100 times such that the war hero wounded 100 times knows that such a man is identical with the war hero wounded 100 times" is true. Yet in all such situations X may fail to know that he himself is the war hero wounded 100 times.5

In a later paper Castaneda refers to the imaginary war hero as "Quintus" and adds that

Quintus studies all accounts of the war hero and discovers that he (the hero) was the only one wounded 100 times. Quintus becomes fascinated by the hero's accomplishments and comes to write the most authoritative biography of the hero. Clearly, for most normal situations, regardless of shifts in the criteria for identifying a person, Quintus knows who the hero was much better than most people, even though Quintus does not know that he himself is the war hero.6

The point Castaneda makes here is very simple. We let 'h' be short for 'the war hero wounded 100 times'. Then although "h knows who h
is" and "h knows who he himself is" are both true, it's false that h knows that he himself is h. Thus, the sentence "There is a person x such that h knows that h=x but he does not know that he himself =x" is true. Therefore, sentences having the same form as (A) should not be treated as uniformly inconsistent, and so KB is incapable of properly handling sentences having the same form as (A).

At the root of the problem, of course, is Hintikka's univocal treatment of "knows who" contexts.

Realizing that Castaneda's War Hero example refutes his claims in "Knowing Oneself" and Other Problems in Epistemic Logic' and that his univocal treatment of "knows who" is indeed at the root of the problem, Hintikka reverses tactics. In his 1967 paper 'Individuals, Possible Worlds, and Epistemic Logic' Hintikka admits that

(1) h knows who h is, and
(2) h knows that he himself is h

require translations different from one another in the formal language after all. He doesn't discuss the sentence

(3) h knows who he himself is,

but clearly its translation can be derived from that given for (2).

What he proposes is the following. Continue to translate sentence (1) as

(4) (Ex)Kh(h=x).

But now translate (2) as

(5) (Ex)(h=x & Kh(h=x)),

and presumably (3) will be ambiguous between (5) and

(6) (Ex)(h=x & Kh(x=x)).
The first thing to notice here is that in KB (4) and (5) are virtually equivalent. This is shown as follows:

(a) \((\exists x)\text{Kh}(h=x)\) \(\subseteq \mathcal{W}\) \(\text{Redog}\)

(b) \(-((\exists x)(h=x & \text{Kh}(h=x)))\) \(\subseteq \mathcal{W}\) \(\text{Redog}\)

(c) \((x)(h \neq x v -\text{Kh}(h=x))\) \(\subseteq \mathcal{W}\) \((b),(\text{C.-E}),(\text{C.-&})\)

(d) \(h \neq h v -\text{Kh}(h=h)\) \(\subseteq \mathcal{W}\) \((a),(c),(109)\)

(e) \(-\text{Kh}(h=h)\) \(\subseteq \mathcal{W}\) \((d),(\text{C.Self.})\)

(f) \(\text{Ph}(h \neq h)\) \(\subseteq \mathcal{W}\) \((e),(\text{C.-K})\)

(g) \(h \neq h\) \(\subseteq \mathcal{W}^{*}\) \((f),(\text{C.P}^{*}), \text{for some compatible w}^{*}\)

(h) \(h = h\) \(\subseteq \mathcal{W}^{*}\) \((\text{C.Self.}=)\)

(a) \((\exists x)(h=x & \text{Kh}(h=x))\) \(\subseteq \mathcal{W}\) \(\text{Redog}\)

(b) \(-((\exists x)\text{Kh}(h=x))\) \(\subseteq \mathcal{W}\)

(c) \(h=b & \text{Kh}(h=b)\) \(\subseteq \mathcal{W}\) \((a),(108), \text{for some b}\)

(d) \((\exists x)\text{Kh}(b=x)\) \(\subseteq \mathcal{W}\) \((a),(108), \text{for some b}\)

(e) \((x)-\text{Kh}(h=x)\) \(\subseteq \mathcal{W}\) \((b),(\text{C.-E})\)

(f) \(-\text{Kh}(h=b)\) \(\subseteq \mathcal{W}\) \((d),(e),(109)\)

(g) \(\text{Kh}(h=b)\) \(\subseteq \mathcal{W}\) \((c),\text{simp.}\)

Now since Castaneda has demonstrated that (1) and (2) are not equivalent, certain changes need to be made in the rules of KB to prevent the inference from (4) to (5), and vice versa. Hintikka recognizes this need but fails to indicate anywhere the specific changes that must be implemented to block these inferences and still make reasonable sense.

It seems to me that the required changes can be made most straightforwardly by a revision of (108) and (109) to:
(108*) If \( (Ex)p \) \& \( w \) and \( x \) occurs free only within the scope of a single epistemic operator \( Kb \) or \( Pb \), then for at least one free individual symbol \( 'a' \) we have \( 'p(a/x)' \& \( w \) and \( '(Ex)Kb(a=x)' \& \( w \); if \( x \) occurs free both within and without the scope of a single epistemic operator \( Kb \) or \( Pb \), then we have \( 'p(a/x)' \& \( w \) and \( '(Ex)(a=x \& Kb(a=x))' \& \( w \).

(109*) If \( '(x)p' \) \& \( w \) and \( x \) occurs free only within the scope of a single epistemic operator \( Kb \) or \( Pb \), then for at least one free individual \( 'a' \), if \( '(Ex)Kb(a=x)' \& \( w \), then \( 'p(a/x)' \& \( w \); if \( x \) occurs free both within and without the scope of a single epistemic operator \( Kb \) or \( Pb \), then if \( '(Ex)(a=x \& Kb(a=x))' \& \( w \), then \( 'p(a/x)' \& \( w \).

Having admitted for the first time that (1) and (2) require different formal treatments, Hintikka in 'Individuals, Possible Worlds, and Epistemic Logic' argues that his change in strategy is necessitated only by Castaneda's employing queer standards in using "knows who" the way he does.

The standards of knowing who presupposed by Casaneda are such that although \( h \) knows who \( h \) is, he cannot point to anyone and say: "That man is \( h \)" or even get hold of \( h \) in other ways. This is precisely what is queer about these standards.... Castaneda is presupposing a standard of knowing who on which one can know who \( h \) is without being able to locate him in the actual world. A standard of this kind is logically possible to have. Whether such kind of 'knowing who' deserves to be called by this name is to my mind very dubious."

It appears that Hintikka wishes to distinguish here between at least two standards (as he calls them) of knowing who. A man can know who \( h \) is and be able to locate him in the actual world, or a man can know who \( h \) is and not be able to locate him in the actual world. What it means to "locate \( h \) in the actual world" is not clear, but we can follow Sleigh's lead and refer to the distinction as between \( h \)'s being "known strictly" versus \( h \)'s being "known weakly".

To say that \( h \) knows who \( h \) is without being able to locate \( h \) in
the actual world, Hintikka feels, one is presupposing a standard of knowing who which "is queer" and hardly deserving of the title 'knowing who'. He therefore seems to regard the truth of (1) in Castaneda's example as based on some sort of less than standard use of English. If we admit this usage of "knows who", then certainly distinct translations of (1) and (2) must be given, and the rules of KB must be changed accordingly. As far as standard usage is concerned (the implication seems to be) KB was right all along, but to accommodate non-standard usage certain changes must be made.

It is ironic that Hintikka himself makes use of the expression "knows who" in this same weak sense. On page 148 of Knowledge and Belief he speaks of our knowing who the teacher of Antisthenes is. And on page 132 we are asked to consider the sentence "a knows who killed Toto", where the reference is to Toto de Brunel, a character of the author Lawrence Durrell. Both characters, it seems clear, would be impossible to track down in the actual world. Each could at best be known weakly by an agent. And so it hardly seems fair to accuse Castaneda of presupposing a standard of "knows who" which is "queer" and hardly deserving of the name.

Castaneda's 1968 paper 'On the Logic of Attributions of Self-Knowledge to Others' is both a reply to Hintikka's 'Individuals, Possible Worlds, and Epistemic Logic' and a discussion of his rather involved theory of indicators alluded to earlier (see footnote 3). In those parts of his paper relevant to his dispute with Hintikka he elaborates further upon his War Hero example and discusses a number of complexities in self-reference brought out by the example. He seems to welcome Hintikka's abandonment of a univocal translation
for (1) and (2) but objects to Hintikka's solution on other grounds.

Consider:

(7) a knows that b knows that c knows that he himself is F and suppose that "he himself" refers to a. According to Hintikka's new system IPE (7) is rendered

(8) (Ex)(a=x & KaKbKc(Fx)).

But if (8) belongs to a model set \( w \), then so do both

(9) a=h & KaKbKc(Fh), and

(10) (Ex)(h=x & KaKbKc(h=x))

for at least one constant 'h'. But (7) can be true even when

(11) a knows that b knows that c knows that h is F is false for every constant term 'h' in the language. It may be false simply because there is no particular way a, b, and c share in common of referring to a. So since (7) does not imply (11) it ought not be the case that (8) virtually imply both (9) and (10) in IPE; hence it looks like there are serious difficulties with IPE.

This brings us to the final stage in the controversy, Hintikka's response to Castaneda in his 1970 paper 'On Attributions of Self-Knowledge'. Regarding Castaneda's allegation that (3) virtually implies (9) and (10) in IPE Hintikka writes:

According to my rule (C.E.), '(Ex)f' can be present in a model set \( \mathcal{M} \) only if, for some individual constant a, 'f(a/x)' is in \( \mathcal{M} \). Therefore (by Castaneda's argument), for some a the presence of '(Ex)f' in entails the presence of 'f(a/x)' in \( \mathcal{M} \). This is an inference from a statement of the form \( p \rightarrow (Ex)q(x) \) to a statement of the form \( (Ex)(p \rightarrow q(x)) \) with '→' representing some sort of entailment. It is clearly fallacious...It is a while since I have seen a competent philosopher being taken in by the old operator-switch swindle as neatly as Castaneda.9
According to Hintikka, Castaneda is mistaken to have supposed that the presence of (8) in a model set necessitates the presence of (9) and (10) in the same set. The relevant quantifier rule (see our (108*)) tells us that if (8) is a member of w, then for at least one free individual symbol -- in this case 'h' -- (9) and (10) belong to w. It does not tell us that for at least one free individual symbol 'h', if (8) belongs to w then so do both (9) and (10). And so, presumably, Castaneda's argument turns on a scope ambiguity, and (8) does not virtually imply (9) and (10).

What is of interest here, I believe, is Hintikka's warning that '→' represents "some sort of entailment". For consider:

(a) If (8) is a member of w, then for at least one free individual symbol 'h', (9) and (10) belong to w

(b) For at least one free individual symbol 'h', if (8) is a member of w, then (9) and (10) belong to w

Here we let the underlined portion in each case indicate what Hintikka's "p" stands for. Clearly there is no problem inferring (b) from (a) if what the "If...then..." represents here is material implication (the symbol 'h' never occurs in (8) and hence not in p). We can agree, however, that if the "If...then..." represents some meta-theoretic entailment, then the inference from (a) to (b) is not clearly legitimate. It is to Hintikka's discredit that he does not explain more fully exactly what's wrong with the fallacious inference and the nature of the entailment in question, but we can grant him the point.

What I fail to see, though, is the relevance of Hintikka's objection. Suppose that Castaneda may not make the move from (a) to (b). Cannot Castaneda's point be cast in terms which bypass this
restriction? Suppose once again that (7) is true. Then we let the formal translation of (7), i.e., (8), belong to the actual world w. If (8) belongs to w, however, then by (a) we have

(c) For at least one free individual symbol 'h', (9) and (10) belong to w.

But isn't this the very difficulty cited earlier, since if (c) is true it follows that

(d) Agents a, b, and c share a common way of referring to a?

As we've stated the argument there is no appeal to (b) either explicitly or implicitly; the point Castaneda makes simply does not depend upon the move from (a) to (b). Hintikka's argument seems to be beside the point, and the serious difficulty Castaneda has uncovered remains.

Be this as it may, Hintikka (in this same paper) nevertheless rejects his proposal to translate (2) as (5) and to make a formal distinction between (4) and (5). He doesn't explain this decision, except to say that an attempt to distinguish formulas like these (which are virtually equivalent in KB) "is doomed to remain fruitless". Instead, Hintikka is now ready to propose a far grander scheme for giving (1) and (2) separate formal readings.

***

We now move to the major area of concern in this chapter, Hintikka's proposal to introduce a second pair of quantifiers to his semantics for knowledge to obtain the expanded system KBC. So far we have seen that on account of Castaneda's War Hero example Hintikka has become convinced that (1) and (2) ought to receive non-equivalent
formal readings. Hintikka's reaction to Castaneda's example ranges from calling it "ingenious" in 'Objects of Knowledge and Belief' (p. 880) to "so artificial as to sound like the parody of a philosophical example" in 'On Attributions of Self-Knowledge' (p. 76). But it is clear that the example has had a monumental impact on Hintikka's approach to formalizing a semantics for knowledge.

The solution to the problem of formally rendering (1) and (2), as Hintikka now proposes, goes in the following way:

It seems to me that the key to Castaneda's interesting example is that in translating (1) and in translating (2) we have to use different quantifiers. To mark this formally, let me use the quantifiers $(\exists x)$, $(\forall x)$ in the intended translation (4) of (1), and let me use the other kinds of quantifiers -- say $(\exists x)$, $(\forall x)$ -- in the translation $(?) (\exists x) Kh(h=x)$ or, possibly $(8) (\exists x)(h=x \& Kh(h=x))$ of (2).

Consider once again the War Hero. We can agree that h (the War Hero) does indeed know who h is, but there are certain things h cannot know about the War Hero.

h cannot know very much about the hero's physical appearance; for otherwise he would recognize himself as the hero, amnesia or no amnesia....When confronted with the question: who is h? he will experience a peculiar embarrassment. He cannot tell where h is now, nor give the questioner any other recipe that would show how to get in touch with the hero. He cannot place the hero into any cognitive Lebenswelt of his own, nor can he help anyone else to place the hero into his sphere of personal acquaintance. Considerations of this kind easily lead us to acknowledge a sense of "knowing who" in which it is false to say in Castaneda's example that h knows who the hero (=h) is....In this sense a knows who b is if and only if a can place b within the range of his personal cognitive experience, in other words, if he is
(in a sense of the word devoid of social overtones) 
acquainted with b.12

We have therefore two sense of "knows who" to distinguish.

Taken in one sense (1) is true; h knows who h (the subject of his 
extensive biography) is. But in another sense, Hintikka urges, (1) 
is false; h cannot place the hero into the range of his own "personal 
cognitive experience". For purposes of convenience let us distinguish 
these two English senses by subscripts. We shall say that "h knows 
who1 h is" is true in Castaneda's example while "h knows who2 h is" 
is not.13 Clearly we are in no position to offer criteria for 
distinguishing these two senses, but we shall nevertheless recognize 
that a distinction is being made and hope to become clearer on what 
Hintikka has in mind in what follows.

Regarding "knows who2" Hintikka writes:

The second sense of "knowing who" is precisely 
the sense expressions of the form $\exists x \forall y (x, y)$ 
have when the quantifier $(\exists x)$ is based on those 
methods of cross-identification which turn on the 
knower's personal cognitive situation...In analogy 
to Russell's "knowledge by acquaintance" we might 
speak here of "individuation by acquaintance".14

And so, as in the case of perception, Hintikka urges that there are 
two methods used by people to individuate objects from world to 
world in those worlds compatible with everything they know. And 
corresponding to these two methods of cross-identification there are 
two sets of quantifiers. One set ranges over world-lines determined 
by one of these methods, and the other ranges over world-lines de-
termined by the other method. Variables bound by $(\exists x)$ and $(x)$ 
range over individuals known1 by the agent, and variables bound by 
$(\exists x)$ and $(\forall x)$ range over individuals known2 by the agent.15 Thus,

(17) h knows who1 h is
is symbolized as (4), and

\[(18) \ h \text{ knows who}_2 \ h \text{ is}\]
is symbolized as

\[(19) (\exists x)Kh(h=x).\]

There is, however, still a crucial question remaining to be answered. Hintikka has argued that (1) is ambiguous between (17) and (18) but that in Castaneda's story (1) is to be read as (17). Hence (4) is that proper formal reading of (1) as far as the story is concerned. Now it would be nice to say that (19) is the proper formal reading of (2). But so far nothing has been said regarding how Hintikka's new concept of cross-identification by acquaintance bears upon the case of the War Hero. To explain the connection Hintikka writes.

...Each use of the first-person pronoun "I" get its reference from the context of utterance -- which for the speaker normally is part of his immediate cognitive environment. Thus, by the same token, the reference of any occurrence of "he himself" is determined by means of the personal cognitive situation of the person in question -- in terms of the frame of reference constituted by his first hand acquaintance of persons, things, places, etc. Thus when "he himself" is identified with someone (presupposing that the phrase is used in the way under consideration), the identification usually makes no sense unless his personal frame of reference is relied on. But this means that \(\exists\) and \(\forall\) are to be used rather than \(\in\) and \(\cup\), yielding precisely the translation \(\mathcal{L}(19)\) for (2).\(^{16}\)

And so everything is as it should be. (1) translates as (4) and (2) translates as (19). The War Hero knows who\( _1\) the War Hero is but doesn't know who\( _2\) the War Hero is. Hence (4) is true and (19) false. Castaneda's problem has been solved, but the solution has required nothing short of a major overhaul of Hintikka's semantics.

There is an interesting sidelight to Hintikka's solution. We
recall that Hintikka understands the sentence '(∃x)(Perceives b=x)' to capture the English direct-object construction "Jones perceives b" in his semantics for perception. Likewise he now urges that a formula such as

\[(20) (\exists x)Ka(b=x)\]

can be understood to capture "approximately" the English direct-object sentence "a knows b". Thus we shall understand "a knows who_2 b is" and "a knows b" as having approximately the same force. Again, as in the case of perception, Hintikka distinguishes here between two types of direct-object constructions. A man a can know b in the sense indicated by (20) (Hintikka calls this a's knowing b "knowingly"), or a can merely know the individual who is in fact b. This weaker sense is captured by the formula

\[(21) (\exists x)(b=x \& Ka(x=x))\]

Presumably, the War Hero knows the War Hero in the weaker, but not the stronger, of these sense (that is, although (19) is false, '∀(∃x)(h=x \& Kh(h=x))' is not).

Let us now press forward and try to become a little bit clearer as to the nature of the two types of quantifiers and the worldlines over which they range. Unfortunately, even Hintikka admits that the distinction between knowing who_1 and knowing who_2 cannot be made perfectly clear. In a number of different places he concedes that distinguishing one method of cross-identification from another in the case of knowledge is less clear than in the case of perception.

What can be said about the distinction between known_1 and known_2 individuals? Suppose we begin by cataloguing everything
Hintikka says about knowing who\_2 in order to focus in a bit more on what he has in mind. 

\( b \) is known\_2 by \( a \) just in case:

(i) \( a \) can place \( b \) in a cognitive Lebenswelt of his own (p. 79);

(ii) \( a \) can place \( b \) within the range of his personal cognitive experience (p. 79);

(iii) \( a \) is acquainted with \( b \) (in a sense of the word devoid of social overtones) (p. 79);

(iv) The reference of '\( b \)' is determined by means of the personal cognitive situation of \( a \) (p. 82);

(v) The reference of '\( b \)' is determined by \( a \)'s firsthand acquaintance of persons, things, places, etc. (p. 81).

It will at the same time be helpful for our purposes to notice what Hintikka has to say about knowing who\_1.

In our bookish and public culture, "knowing who" is likely to be interpreted in terms that are independent of the history and situation of the particular person or persons in question.\_18

The War Hero knows who\_1 the War Hero is solely through books, old newspaper accounts, war documents, and the like, and not through anything he can remember personally witnessing.

On the basis of all of this it seems that while Hintikka's distinction may be impossible to specify in a general form, there are some instances of "knows who" that are reasonably easy to identify vis-à-vis the distinction. It's reasonably clear, for example, that knows who\_1 is the sense in which we know who Thomas Edison, Moses, or the Prime Minister is. And in the sentence, "Jones knows who the girl sitting on his lap is", there is little doubt that a firsthand acquaintance has developed between Jones and the girl and that knows who\_2 is the sense in which Jones knows who the girl is.
In probably the majority of cases, however, it's a toss-up to determine which sense of "knows who" is being employed, given what Hintikka has said. I know who the Mayor is, but even though most of my knowledge of the Mayor is based on newspaper, television, and general heresay, I have actually seen him in person. Perhaps I have seen him (from a distance) crown the winning contestant at a beauty pageant. Does this count as first-hand acquaintance? Can I place the Mayor within the range of my "personal cognitive experience"? There seems to be no clear guidelines to decide.

It does seem safe to say, on the other hand, that known₁ and known₂ individuals stand in no subset/superset relationship. In Castaneda's example h is known₂ to h; that is, 'h' picks out a world-line in such a way that makes (4) true and (19) false. It seems too as though the opposite can take place, as in the case perhaps of a man who rides the same bus every morning with a but with whom a has never spoken. Although a knows next to nothing about the man, there is a sense in which a does know who the man is. This sense would seem to be Hintikka's knowing who₂.

Another safe assumption to make concerning the distinction is that formulas of the form

\[(22) \ (Ex)(\exists y)Ka(b=x & x=y)\]

ought to be defensible in KBC. That is, Hintikka's distinction looks as though it allows for a common ground between that which is known₁ and known₂ by an agent. If I were a personal friend of the Mayor I might know who₂ he is on the basis of this friendship, but I might still know who₁ he is on the basis of reading newspaper
articles. Thus Hintikka's distinction may not be mutually exclusive. Conceivably an individual b can be both known\textsubscript{1} to a and known\textsubscript{2} to a at the same time. World-lines formed by different criteria of cross-identification would turn out to coincide in such a case.

In fact, it is tempting to suppose that "sudden recognition" often consists in world-lines established by different criteria suddenly coming together and forming a single world-line. Our man a might discover that the man who rides his bus every morning is the General Manager of the hometown baseball team. Up until now the General Manager has been known\textsubscript{1} by a and the man at the busstop has been known\textsubscript{2} by a, but a failed to know that the man was the General Manager. Hence the respective world-lines until now diverged and only happened to intersect in the actual world. But by virtue of a's discovery it would seem as though '(\text{Ex})(\exists y)Ka(x=the General Manager & y=the man at the busstop & x=y)' is now true. This same sort of phenomenon is noted by both Hintikka and Thomason in the case of perception. A perceptual world-line often turns into a rigid world-line when the agent suddenly recognizes (perceptually) something he's been perceiving, and this yields something of the form '(\text{Ex})(\exists y)(\text{Perceives}(b=x & x=y))'.

Like Hintikka's earlier systems, KBC has a restricted range feature. In KB variables within the scope of 'Ka' and bound from the outside by '(\text{Ex})' or '(x)' range over rigid world lines which pick up individuals\textsubscript{1} in the actual world only if a knows who or what they are. Speaking loosely, we might say that the quantifiers range over individuals only if they are known by a. Now having distinguished formally two senses of knowing who and two sorts of quantifiers
corresponding to these two sense, it turns out that in KBC restricted range is a double-barrelled feature. While the standard quantifiers range over individuals known_1 by a, '(∃x)' and '(∀x)' range over individuals known_2 by a.

Rules governing the new quantifiers are not stated by Hintikka, but for the most part there would seem to be nothing difficult in offering them. Suppose that 'b' is a free individual symbol, and suppose that 'p' contains exactly one occurrence of 'Ka'. Then

\[(108∃) \quad (\exists x)p(x) \in w \quad (109∃) \quad (\forall x)p(x) \in w \]
\[(\exists x)Ka(b=x) \in w \quad (\forall x)Ka(b=x) \in w \]
\[p(b/x) \in w \quad p(b/x) \in w \]

will be the desired quantifier rules. Rules for more complicated constructions can easily be formulated on the basis of general guidelines found in Hintikka's paper 'Existential and Uniqueness Presuppositions'.

Hintikka is also not clear as to whether it is supposed to be the case that if an individual b is known_2 by a, it follows that b exists? This is a question of considerable interest; ought we adopt a rule

\[(C.∃K=) \quad (\exists x)Ka(b=x) \in w \quad (\forall x)Ka(b=x) \in w \]
\[(Ex)(b=x) \in w \quad (Ex)(b=x) \in w \]

in KBC? Again, Hintikka offers no advice here; he has, of course, repeatedly chosen to reject the corresponding rule (C.EK=) for ordinary quantifiers. But there seem to be good grounds to accept (C.∃K=) even if (C.EK=) is not accepted.

Though Hintikka never explains precisely why he dropped (C.EK=) from his original system KB, evidence suggests that the major reason was his desire to consistently symbolize such sentences as "a
knows who murdered Toto de Brunel" and "a knows who puts toys in children's stockings". In his semantics for perception, for example, the move from '(Ex)(Perceives b=x)' to '(Ex)(b=x)' is ruled out precisely to allow one to say that Jones perceives who b is without implying that b exists. With (C.∃K=), however, the situation seems different. It's extremely hard to imagine how something non-existent could be known² by an agent; how could a "firsthand acquaintance" be the basis for a's knowing who something is if it fails to exist? We'll leave open the question whether KB has the rule (C.∃K=) since Hintikka has said nothing on the subject; we'll assume only that (C.EX=) does not seem to be a rule Hintikka accepts. But (C.∃K=) certainly does seem reasonable.

We have already made mention of Hintikka's rule (C.Ind.=) in connection with Thomason's semantics for 'sees that'. There we identified it as a no-splitting condition for world-lines over which variables bound by '(Ex)' and '(x)' range. It goes as follows:

\[
\text{(C.Ind.=)} \quad \begin{align*}
(Ex)(b=x & \land Ka(b=x)) & \in w \\
(Ex)(c=x & \land Ka(c=x)) & \in w \\
K_a(b=c) & \in w.
\end{align*}
\]

If 'b' and 'c' pick out world-lines over which variables bound by '(Ex)' range, and if these world-lines intersect in the actual world, then they must intersect everywhere.

Hintikka's original semantics for knowledge failed to have a no-splitting condition such as (C.Ind.=), but it was shown by Sleigh in his 1967 paper 'On Quantifying Into Epistemic Contexts' that the absence of a no-splitting condition created significant problems in KB. There we are asked to consider:
(a) Everyone known to a is known by a to be F
(b) Everyone is known by a to be F
(c) There is someone known to a whom a does not know to be F
(d) (x)KaFx
(e) (x)(Ey)(x=y & KaFy)
(f) (Ex)-KaFx.

Here (a)-(c) are the English readings for (d)-(f), respectively.
Therefore we would expect that since (b) implies (a), (e) should imply (d); and since (b) and (c) are inconsistent, we should expect that (e) and (f) are mutually indefensible. However, without (C.Ind.=) or something equivalent it cannot be proven in KB that (e) implies (d) or that (e) and (f) are mutually indefensible. Therefore, it is absolutely essential that (C.Ind.=) or something equivalent be made a rule; realizing this, Hintikka employed (C.Ind.=) in his 1967 revised system IPE. Interestingly, there is no similar argument to show that the comparable no-merging condition:

\[
\begin{align*}
\text{(C.Ind.$\neq$)} & \quad (\text{Ex}(b=x & Ka(b=x)) \in w \\
& \quad (\text{Ex})(c=x & Ka(c=x)) \in w \\
& \quad b \neq c \\
& \quad Ka(b \neq c) \in w
\end{align*}
\]

needs to be a rule in Hintikka's system, and consequently it is never adopted.

Where does this leave KBC on the question of splitting and merging? Unfortunately, there is very little to go on here; nothing is clearly spelled out in papers written since adopting KBC. In 'The Semantics of Modal Notions and the Indeterminacy of Ontology', for example, he writes that a no-splitting condition is a limitation which "seems in order", but he fails to specify which kind of world-lines
he has in mind. And in 'Objects of Knowledge and Belief' he remarks that we have "tacitely presupposed" that world-lines do not split, but there is no clear indication whether he has presupposed no-splitting relative to both kinds of quantifiers or to one. These passages suggest that KBC has either (C.Ind.=) or

\[
\begin{align*}
(C.\text{Ind.} \exists) & \quad (\exists x)(b=x & Ka(b=x)) & \in W \\
(C.\text{Ind.} \forall) & \quad (\exists x)(c=x & Ka(c=x)) & \in W \\
\text{b=c} & \quad Ka(b=c) & \in W,
\end{align*}
\]

but it's not clear that KBC has both.

More important than the question whether Hintikka actually intends to employ these rules in KBC is the question whether he ought to. In the final section of this chapter I argue that by a modified version of Sleigh's argument it is necessary for Hintikka to accept both rules. Failure to do so will result in the failure to capture certain vital inferences in KBC.

Before moving on, it is worth pointing out that there is a further -- and more interesting -- reason why (C.Ind.=) should have been employed in KB. According to Hintikka, quantifiers in KB which bind variables occurring free in epistemic contexts range over what he calls "genuine individuals". Using terminology we developed in Chapter One, we can say that an individual sub 2 b is a "genuine individual", as Hintikka uses the term, just in case there is an x such that x is an individual sub 1 and x is the value of b at every world. In other words, an individual sub 2 is genuine just in case it's a constant function.

Now consider '(Ex)(b=x & Ka(b=x))' and '(Ex)(c=x & Ka(c=x))'. In each formula 'x' occurs within the scope of an epistemic operator
and is bound by '(Ex)' from the outside. Therefore, the first formula asserts that \( b \) is a genuine individual, and the second that \( c \) is a genuine individual. Thus, 'b' and 'c' each denote a constant function defined over worlds. And so if we suppose in addition the condition
\[ b = c \in W, \text{i.e.,} \]
that 'b' and 'c' happen to assign to \( w \) the same individual, it follows from everything we have said that 'Ka(b=c)\( \in w \) is true. Constant functions sharing the same value somewhere share the same value everywhere. Therefore, if we take seriously the notion of Hintikka's "genuine individuals", we should be able to prove 'Ka(b=c)\( \in w \) given that 'b' and 'c' denote genuine individuals and that 'b=c\( \in w \). Without something like (C.Ind.=), however, this cannot be done.

We might put this argument as follows. Choose an intended possible worlds interpretation \( \langle D, W, R \rangle \) for Hintikka's system in such a way that '(Ex)(b=x & Ka(b=x))' is true in a world \( w \in W \) if and only if there is some element \( d \in D \) such that in every world \( w' \) such that \( \text{Rel} \) the term 'b' designates \( d \) in \( w' \). Then '((Ex)(b=x & Ka(b=x)) & (Ex)(c=x & Ka(c=x)) & b=c) \rightarrow Ka(b=c)' will be true in every world \( w \) relative to \( \langle D, W, R \rangle \) even when it cannot be proven in KB. And so the rules of KB are not powerful enough to prove formulas true in every world in the intended model. The idea, of course, is that (relative to \( \langle D, W, R \rangle \)) if both 'b' and 'c' each denote the same element in every compatible world, then if 'b=c' is true in any world it's true in every compatible world.
In this section we consider difficulties which beset Hintikka's expanded system KBC.

1. **KBC cannot be expanded to a workable system of knowledge + belief.**

One of the most attractive features of KB when it first appeared was the obvious way in which it could be expanded to a system containing both the epistemic operator 'Ka' and the doxastic operator 'Ba'. In *Knowledge and Belief* Hintikka talks informally of such a system and discusses some interesting applications it is seen to have ("Moore's Paradox", "Thinking that one might be mistaken", etc.). We might similarly try now expanding KBC to a system containing the belief operator 'Ba'.

The first thing to notice about such a system -- let us call it KBCB -- is that while there are two methods by which objects are cross-identified relative to what is known, there is only one method according to which world-lines are formed across doxastic alternatives (worlds compatible to what is believed). According to Hintikka, there is no parallel in belief to the method of "contextual" cross-identification according to which individuals are known by an agent. A single set of quantifiers suffices to exhaust what Hintikka feels to be the complexities in the logic of belief.

What happens when we actually set up a system in which variables in epistemic contexts may be bound by two sorts of quantifiers, but variables in doxastic contexts may be bound only by '(Ex)' and '(x)'? Trouble as far as I can see. Consider:

(33) Someone is known by a to be F

(34) Someone is believed by a to be F.

It's reasonable to suppose that (33) entails (34), but although (33)
is translated by either (Ex)KaFx or (\exists x)KaFx, (34) is unambiguous and is translated (Ex)BaFx’. To make sure that the entailment of (34) by (33) is maintained in KBCB, therefore, it would seem that both

\[
\frac{(Ex)KaFx}{(Ex)BaFx} \quad \varepsilon \quad w \quad \frac{(\exists x)KaFx}{(Ex)BaFx} \quad \varepsilon \quad w
\]

ought to be provable in the system. Now consider

(35) Someone is known\textsubscript{1} by \(a\) to be \(F\) and known\textsubscript{2} by \(a\) to be \(G\),

which is symbolized

(36) (Ex)(\exists y)(x=y & KaFx & KaGy).

Given the derivations above it can be proven that (36) virtually implies

(37) (Ex)(Ey)(x=y & BaFx & BaGy).

Now (37) reads

(38) Someone is both believed by \(a\) to be \(F\) and believed by \(a\) to be \(G\),

and this seems to raise problems. For suppose \(a\) is the War Hero; then (35) is true if ‘\(F\)’ is short for 'has been wounded 100 times' and '\(G\)’ is short for 'is the War Hero’s biographer’. But it seems wrong to infer from the truth of (35) that one and the same person is both believed to be \(F\) and believed to be \(G\). In short, KBCB ought to formally reflect the entailment of (34) by (33), but with two senses of "known" and one sense of "believed" we run into difficulty. The entailment of (34) by (33) either fails for one sense of "known" or we must countenance what appears on the surface to be a bad inference (from (35) to (38)).

At this point I'm not entirely convinced that the inference from (35) to (38) is bad. Nevertheless, this is not the end of the
Gail Stine argues in 'Hintikka on Quantification and Belief' that any Hintikka-type semantics for belief must contain the rule:

\[(\exists x)(b=x \land Ba(b=x)) \in w \]
\[(\exists x)(c=x \land Ba(c=x)) \in w \]
\[b=c \Rightarrow Ba(b=c) \in w.\]

With "Stine's Rule" (36) virtually implies not only (37) but

\[(39) (\exists x)Ba(Fx \land Gx).\]

Now this is clearly problematic. For it follows for some free singular term 'd' that 'Ba(Fd \land Gd)' is true; that is, a believes that the same person is both F and G. In particular, whenever someone is known₁ by a to be b and known₂ by a to be c, then if b=c it follows that a believes that b=c. Our man at the busstop can't fail to believe that the General Manager is identical with the man he sees every morning boarding the bus. And the War Hero cannot fail to believe that he himself is the War Hero.

It would be easy to avoid this difficulty by deciding that Stine's Rule ought not be incorporated in KBCB. This would certainly be a heavy-handed way of dealing with an unwanted inference; but, more importantly, it would raise up a serious problem put to rest by Stine's adopting her rule. Consider the belief version of the argument of Sleigh's we looked at earlier.

\[(40) \text{Everyone is believed by a to be F}\]
\[(41) \text{Everyone of whose identity a has a true opinion is believed by a to be F}\]
\[(42) \text{There exists someone of whose identity a has a true opinion and of whom a does not believe that he's F}\]
\[(43) (x)(Ey)(x=y \land BaFy)\]
\[(44) (x)BaFx\]
\[(45) (Ex)-BaFx.\]
Formulas (43)-(45) are formal renderings of (40)-(42), respectively. Therefore, (i) One would expect that since (40) entails (41), (43) should virtually imply (44), and (ii) Since (40) and (42) are logically inconsistent, the conjunction of (43) and (45) should be indefensible. However, (43) does not virtually imply (44), and (43) and (45) can in fact be imbedded in the same model set. Hence there is something wrong with the semantics for belief originally proposed by Hintikka.

By adding to the semantics Stine's Rule, however, both difficulties can be avoided. We prove that (43) virtually implies (44) in B(alt), the system Stine advocates.

\[\begin{align*}
(a) & \quad (x)(Ey)(x=y & BaFy) \quad \in w \\
(b) & \quad -(x)BaFx \quad \in w \\
(c) & \quad (Ex)-BaFx \quad \in w \quad (b), (C.-U) \\
(d) & \quad (Ex)(b=x & Ba(b=x)) \quad \in w \\
(e) & \quad -BaFb \quad \in w \\
(f) & \quad (Ex)(b=x) \quad \in w \quad (d) \\
(g) & \quad (Ey)(b=y & BaFy) \quad \in w \quad (a), (f) \\
(h) & \quad (Ey)(c=y & Ba(c=y)) \quad \in w \quad (g), for some c \\
(i) & \quad b=c & BaFc \quad \in w \\
(j) & \quad b=c \quad \in w \quad (i), Simp. \\
(k) & \quad Ba(b=c) \quad \in w \quad (d), (h), (j) \\
(l) & \quad BaFc \quad \in w \quad (i), Simp.
\end{align*}\]

Clearly, however, (e), (k), and (l) entail a contradiction, that in some alternative \(w^*\) of \(w\) both 'Fb' and '¬Fb' belong to \(w^*\). In this proof (d), (e), (g), and (h) are obtained via Stine's rule (103)''\(\), (f) comes by way of her rule "II."\(\), and the key step (j), upon which
the whole proof depends, is inferred on the basis of "Stine's Rule".

The choice is clear. Either KBCB can incorporate Stine's Rule and avoid this difficulty at the expense of enabling (36) to virtually imply (39); or KBCB can reject Stine's Rule, thereby blocking the move from (36) to (39) but at the same time blocking the move from (43) to (44) and the proof that (43) and (45) are mutually indefensible. Either way there's big trouble; either way it appears that KBCB is not a workable system. That KBC cannot be expanded to a workable system of knowledge + belief seems to be a serious disadvantage of KBC itself.23

2. Problems with (C.Ind.=) and (C.Ind^) in KBC. Earlier we remarked that a variant of Sleigh's first argument in 'On Quantifying into Epistemic Contexts' could be used to show that both (C.Ind.=) and (C.Ind^) must be rules in KBC.

\( (40') \) Everyone is known by \( a \) to be F

\( (41') \) Everyone who exists and is known by \( a \) is known by \( a \) to be F

\( (42') \) There exists someone known by \( a \) who is not known by \( a \) to be F

\( (43') \) \((x)(Ey)(x=y & KaFy)\)

\( (44') \) \((x)(((Ey)(x=y) & Ka(x=x)) \rightarrow KaFx)\)

\( (45') \) \((Ex)(Ey)(x=y & \neg KaFx)\).

As above, \( (40')-(42') \) are formally rendered by \( (43')-(45') \), respectively. It should be the case in KBC, therefore, that \( (43') \) virtually imply \( (44') \) and that \( (43') \) and \( (45') \) be mutually indefensible. However, this may be proven in KBC only with the use of (C.Ind.=) or something equivalent to it. And a perfectly parallel argument shows that (C.Ind^) must be likewise present in KBC.
We must therefore assume that both conditions are rules in the new system; world-lines of neither type are allowed to split.

Formally speaking, there is nothing problematic in employing both (C.Ind.=) and (C.Ind\(\exists\)). But I believe that a serious question arises whether these principles are too strong from the point of view of ordinary knowledge. In 'Existential and Uniqueness Presuppositions' Hintikka admits that what appear to be counter-examples have been raised against (C.Ind.=), but all such examples, he maintains, turn on the use of different criteria for knowing who. Thus, for example, Castaneda's War Hero case can be viewed as a counter-example to (C.Ind.=): The War Hero knows who the War Hero is, he knows who he himself is, but he fails to know that he himself is the War Hero. This example, however, depends upon different criteria for knowing who; two different methods of cross-identification are being employed, and in KBC these methods are formally distinguished. Hence the War Hero example is not a counter-example to (C.Ind.=) in KBC.

In light of this, consider the following two examples. First, Jones is an ardent basketball fan. He knows who Walt Hazzard is, and he knows who Abdul Rahman is; with respect to each he knows a wealth of statistics. Jones has, however, never seen the man in person. And as much as Jones knows about the man, Jones fails to know that Walt Hazzard is Abdul Rahman. Second, Jones knows who his father is. Jones also knows who the winner of the masquerade contest is (it's the gorilla, and Jones has just congratulated him). Yet Jones fails to know that it is his father who has just won the contest.
In the first example Jones knows who both Hazzard and Rahman are, but notice that he cannot know who\_2 either of them are; neither of them is known\_2 to Jones. Therefore, either they are both known\_1 to Jones and we have a counter-example to \((C.\text{Ind}.=)\), or our example turns on employing "different criteria" for knowing who, one of which has so far been unaccounted for by Hintikka.

And in the second example both Jones's father and the winner of the contest are known to Jones, but neither of them is known\_1 to Jones. Therefore either both are known\_2 to Jones and we have a counter-example to \((C.\text{Ind}.\exists)\) in KBC, or our example turns on employing "different criteria" for knowing who, one of which has so far been unaccounted for by Hintikka.

Thus, either we have counter-examples to both \((C.\text{Ind}.=)\) and \((C.\text{Ind}.\exists)\) in KBC, or there are methods of cross-identification yet to be worked into the formal semantics. If the latter is true, then there will no doubt have to be still more quantifiers added to the system to capture all that needs to be said. Either way the situation does not look encouraging for KBC. The very sort of trouble the two sets of quantifiers were supposed to avoid has not been avoided at all; making formal distinctions between different senses of "knowing who" by introducing quantifiers does not seem to be a promising line to pursue.

3. **Rigid and non-rigid world-lines in KBC.** Much of the appeal generated by Hintikka's original system KB lay in the obvious way in which Hintikka's model sets could be thought of in terms of possible worlds over which could be defined a compatibility relation determined by what is known by a given agent. We looked earlier at
an account of such an intuitive interpretation of \( \text{KB} \), according to which a possible worlds semantics \( \langle W, D, R \rangle \) can be identified as the intended model of the formulas of \( \text{KB} \).

Central to the specification of \( \langle W, D, R \rangle \) was the set of truth conditions for formulas of the form \( (\exists x)(b=x & \text{Ka}(b=x)) \): true in a world \( w \in W \) just in case there is an element \( d \in D \) such that in every world \( w^* \) such that \( R w^* \), \( \not\models (b, w^*) = d \). An agent \( a \) knows who someone \( b \) is only if in every state of affairs compatible with everything \( a \) knows, \( 'b' \) picks out the same person. Such a condition has intuitive appeal. It would seem wrong to say that \( a \) knows who \( b \) is if, so far as \( a \)'s knowledge is concerned, \( b \) could be any number of people. To really know who the masked man is is to know with respect to some one particular person that he's the masked man. In Hintikka's terms, the masked man must be a "genuine individual".

It is unfortunate, I believe, that such an intuitively appealing interpretation of Hintikka's system is forfeited by the move to \( \text{KBC} \). It is not the presence of two sets of quantifiers in the formal language which necessitates the abandonment of this simple interpretation. Rather, it is this: The truth-conditions for \( (\exists x)\text{Ka}(b=x) \) and \( (\exists x)\text{Ka}(b=x) \) in \( \langle W, D, R \rangle \) are non-equivalent; hence it must be that \( 'b' \) in either \( (\exists x)\text{Ka}(b=x) \) or \( (\exists x)\text{Ka}(b=x) \) does not pick out a rigid world-line. Hence the world-lines associated with either known or known individuals (or both) are wobbly. In general, it is no longer the case that if \( a \) knows who \( b \) is then \( b \) is a genuine individual, an individual which remains invariant over all worlds compatible with what \( a \) knows.

Perhaps (as in his semantics for perception) Hintikka has this
in mind: that both known₁ and known₂ individuals are invariant over the set of worlds but invariant in different respects. That is, according to two different criteria of identity, the elements chosen by the respective world-lines are identical with one another. Hence if b is known₁ by a, b picks out the same individual in each world according to one criteria of sameness, and if b is known₂ by a, b picks out the same individual in each world according to a different criteria of sameness.

If this is the case then clearly \( \langle W, D, R \rangle \) is unable to model the formulas of KBC. Something far more sophisticated and far less intuitive is required, and what that might be I have no idea.
FOOTNOTES

1 Hintikka, Knowledge and Belief, p. 149, n. 9.

2 Castaneda, Review of Knowledge and Belief, p. 134.

3 In 'On the Logic of Self-Knowledge' Castaneda considers the sentence "a knows that b knows that he knows that he is tall" and finds that the "c" operator alone can't make all the necessary distinctions. In later papers Castaneda proposes a general system to handle cases of arbitrary complexity, but such is beyond the scope of our discussion.

4 Hintikka, 'Knowing Oneself', pp. 10-12.

5 Castaneda, 'On the Logic of Self-Knowledge', pp. 11-12.

6 Castaneda, 'On the Logic of the Attribution of Self-Knowledge to Others', p. 446.

7 Hintikka, Individuals, Possible Worlds, and Epistemic Logic', pp. 52-53.


9 Hintikka, 'On Attributions of Self-Knowledge', p. 75.

10 Ibid., p. 77, n. 6.

11 Ibid., p. 78. By coincidence Hintikka's (1), (2), and (4) correspond exactly with our sentences of those numbers.

12 Ibid., pp. 78-79.

13 There is no reason to believe that this distinction is the same distinction as the one made earlier between knowing who strictly and knowing who weakly.

14 Hintikka, 'On Attributions of Self-Knowledge', p. 79.

15 We shall say that x is known₁ (known₂) by y just in case y knows who₁ (knows who₂) x is.


17 See for example, 'Different Constructions in Terms of Basic Epistemological Terms', p. 119.


19 See Hintikka's brief paper 'Reply'.
This point is made by Sleigh in his 1967 paper also, but his explanation (involving talk of the "indiscernibility of identicals") I find somewhat misleading.

See 'Objects of Knowledge and Belief', p. 801, and 'Knowledge by Acquaintance', p. 66.

The same proof demonstrates the mutual indefensibility of (43) and (45).

There are other problems with KBCB. There seems to be no way, for example, to consistently symbolize sentences like, "There is someone x such that h believes that h=x but h does not believe that he himself is x", the belief version of Castaneda's (A).

The \( \phi \) function is explained by Hintikka in 'Semantics for Propositional Attitudes', p. 92, as mapping pairs of constant terms and worlds to members of a large domain.
CHAPTER IV

There has been no single issue discussed in connection with Hintikka's semantics for propositional attitudes more than the so-called Restricted Range feature (though not always under this name). Already we have had occasion several times to make reference to this feature, but so far we have not addressed ourselves to the question whether this feature is avoidable in a Hintikka-type semantics. In this chapter and the next we shall attempt to answer this question.

We shall proceed as follows: (i) Discuss what is meant by "Restricted Range" and why Hintikka holds that the feature is unavoidable in his systems; (ii) Take notice of several criticisms of this feature which have appeared in print; (iii) Present variants of Hintikka's system KB proposed by Føllesdal, Tienson, and Sleigh which lack the feature; and (iv) Point out disadvantages in the approach taken by each. Then in the next chapter we shall present a variant of KB minus Restricted Range which avoids these disadvantages. It shall be concluded that the Restricted Range feature is avoidable in Hintikka-type systems.

In what follows, unless specifically stated to the contrary, our remarks about Hintikka will be confined to his original system KB.

***

There are doubtless a great many things one might mean by speaking of "Restricted Range". Let us set the record straight by explaining precisely how the term has been employed in connection with Hintikka-type semantics for propositional attitudes. Informally
the Restricted Range feature comes to the following. Consider an agent \( a \) and a world \( w \). In describing \( a \)'s knowledge relative to \( w \) quantifiers binding variables which occur free inside the scope of a single epistemic operator range over world-lines which do not pick up every individual in the domain of \( w \). Rather, they range over world-lines which pick up only that subset of individuals in \( w \)'s domain whose members are known to \( a \). Speaking loosely, we might say that the quantifiers are restricted to that subset. Thus '\((x)KaFx\)' is not read "Everyone is known by \( a \) to be \( F \)" but "Everyone who is known to \( a \) is known by \( a \) to be \( F \)" (supposing for the sake of simplicity throughout that we're talking about domains of persons). In this way '\((x)KaFx\)' and '\((x)(Ka(x=x)\rightarrow KaFx)\)' are virtually equivalent.

A formal definition of "Restricted Range" can be based in part on this virtual equivalence. We shall say that a Hintikka-type semantics for knowledge possesses the Restricted Range feature if and only if: (i) All instances of '\((x)(Ka)^{n}Fx \leftrightarrow (x)((Ka)^{n}(x=x)\rightarrow (Ka)^{n}Fx)\)' are valid, where '\((Ka)^{n}\)' is short for \( n \) iterated occurrences of 'Ka', for any positive integer \( n \), and (ii)

\[
\frac{(x)KaFx}{(x)Fx} \in w
\]

is not valid. To extend this definition to systems for other propositional attitudes, replace 'Ka' throughout by 'Perceives', 'Ba', and the like; and where reflexivity fails in a system (such as with 'Ba') replace (iii) by

\[
\frac{(x)(Fx \& BaFx)}{(x)(Fx)} \in w, \quad \frac{(x)(Fx \& Perceives Fx)}{(x)(Fx)} \in w
\]

and so forth.
Let us now examine Hintikka's reasons for holding that Restricted Range is unavoidable in his systems. Those who have considered Restricted Range to be a drawback to Hintikka's approach and have looked for ways to eliminate it have been confronted by Hintikka with a well-known "proof" that Restricted Range cannot be dropped from KB.

Castaneda thinks of the restriction as a dodge for meeting some of Quine's criticisms of quantified modal logic. However, the restriction is not an ad hoc device calculated to meet certain specific criticisms. It can be shown that it has to be adopted by everyone who countenances quantification into a context governed by "Ka" or "Pa" in their normal sense and who accepts the normal meaning of our epistemic notions and of logical connectives as codified by suitable semantic conditions.\(^1\)

The proof goes as follows.

We begin by considering formulas in which quantifiers bind variables occurring free inside the scope of a single epistemic operator. For simplicity consider 

\[ (x)p(x) \] and 

\[ (\exists x)p(x) \].

To determine whether the quantifiers range over a mere subset of existent individuals it must be determined how rules governing these quantifiers are going to be drawn up. In particular, under what conditions may we instantiate and generalize upon 

\[ (x)p(x) \] and 

\[ (\exists x)p(x) \]?

Suppose we let our quantifier rules take the following form:

\[
\begin{align*}
\text{(C.U.)' } & \quad (x)p(x/b) \in \mathcal{W} & \quad (C.E.)' & \quad (\exists x)p(x/b) \in \mathcal{W} \\
& \frac{Q(d) \in \mathcal{W}}{p(d/b) \in \mathcal{W}} & \frac{Q(a) \in \mathcal{W}}{p(a/b) \in \mathcal{W}},
\end{align*}
\]

where 'a' is a new constant and 'Q' is some yet unspecified formula of one free variable. It can easily be proven, Hintikka maintains, that 

\[ A(b) \rightarrow (\exists x)Ka(b=x) \]

is self-sustaining and hence that 

\[ (\exists x)Ka(b=x) \] or something logically equivalent must serve as the
instantiating and generalizing conditions for \((x)p(x)\)' and \'(Ex)p(x)'.

This is easily demonstrated:

\[
\begin{align*}
(a) & \quad A(b) \quad \in \quad w \} \\
(b) & \quad -(Ex)Ka(b=x) \quad \in \quad w \} \quad \text{Redog} \\
(c) & \quad (x)Pa(b\neq x) \quad \in \quad w \} \quad (b),(C.-E.),(C.-K) \\
(d) & \quad Pa(b\neq b) \quad \in \quad w \} \quad (c),(a),(C.U.)' \\
(e) & \quad b\neq b \quad \in \quad w* \quad (d),(C.P*), \text{ for some } w* \text{ such that } R_{ww*} \\
(f) & \quad b=b \quad \in \quad w* \quad (C.self.\text{-}=) \\
(a) & \quad (Ex)Ka(b=x) \quad \in \quad w \} \quad \text{Redog} \\
(b) & \quad -Q(b) \quad \in \quad w \} \\
(c) & \quad Ka(b=d) \quad \in \quad w \} \quad (a),(C.E.)' \\
(d) & \quad Q(d) \quad \in \quad w \} \quad (c),(d),(97) \\
(e) & \quad Q(b) \quad \in \quad w \} \quad (c),(d),(97)
\end{align*}
\]

Supposing, then, that we let the quantifier rules of KB be \((C.E.)'\) and \((C.U.)'\), we are saddled with the Restricted Range feature. According to Hintikka:

Hence we may conclude that, for the kind of 'p' we assumed we were dealing with, "(Ex)Ka(b=x)" has to be assumed as-a-prerequisite for quantifying with respect to the term 'b' in 'p'....Thus the restriction of the substitution-values of bound variables to such individual constants b as satisfy this condition is not just an 'ingenious device' but a necessity which I do not see any chance of escaping.\(^2\)

Therefore not only does KB have the Restricted Range feature, but KB could not have lacked it.

Criticisms of various sorts have been levelled at Hintikka's
proof, but for the time being we shall accept it at face value. By eventually arguing that there exists a plausible modification of KB which lacks Restricted Range we shall in effect suggest that Hintikka's proof is defective. At that point we shall argue that the proof is defective and try to identify at exactly which point it breaks down. But before we do any of that, we turn to criticisms of the Restricted Range feature.

* * *

Perhaps most interesting among the criticisms of the Restricted Range feature which have appeared in print are those raised by Castaneda. In his Review of Knowledge and Belief Castaneda charges that the English sentence

(1) There exists an object of which a does not know that it exists

cannot, as a result of Restricted Range, be consistently rendered in the language of KB. For consider:

(2) (Ex)-Ka(Ey)(x=y).

Since the quantifier ranges over individuals known to a, (2) virtually implies '(Ex)Ka(b=x)' and '-Ka(Ey)(b=y)', for some constant 'b'; but since '(Ex)Ka(b=x)' virtually implies 'Ka(Ex)(b=x)', (2) virtually implies a contradiction. Hence (2) cannot serve as a consistent formal rendition of (1).

Hintikka replies in "Knowing Oneself" and Other Problems in Epistemic Logic' that although Castaneda is correct in pointing out that (2) is indefensible in KB, he (Hintikka) has meanwhile come to decide on independent grounds that the rule
ought to be given up. These independent grounds, as we have pointed out elsewhere, apparently come to this: that \((C.EK.=*\) does not allow one to symbolize such sentences as "a knows who puts toys in children's stockings" and is therefore too restrictive. Giving up \((C.EK.=*\) turns out to be especially convenient for answering Castaneda's objection, however, in that \((2)\) no longer virtually implies a contradiction.

In his 1969 paper 'On the Logic of the Ontological Argument' Hintikka goes a step further and points out that even if \((2)\) turns out to be indefensible in \(KB\), this does not show that \((1)\) has no formal counterpart in \(KB\). For the formula

\[
(3) \quad (Ex)(y)(x=y \rightarrow -Ka(Ez)(y=z))
\]

is both a formalization of \((1)\) and perfectly consistent in \(KB\) (with or without \((C.EK.=*\)).

Much the same can be said of the English sentence

\[
(4) \quad \text{There exists someone of which } a \text{ doesn't know who he is}
\]

(a slight variant of which is charged by Castaneda in 'On the Logic of Self-Knowledge' to be untranslatable in the language of \(KB\)). If quantifiers range over persons known to \(a\), how can it be said of someone that he fails to be known to \(a\)? That is, the formula

\[
(5) \quad (Ex)-(Ey)Ka(x=y)
\]

is outright inconsistent in \(KB\). And the presence or absence of \((C.EK.=*\) is irrelevant to the indefensibility of \((5)\). How then do we capture \((4)\) in formal terms? The answer is easy given Hintikka's proposal to translate \((1)\) as \((3)\); we simply translate \((4)\) as

\[
(6) \quad (Ex)(y)(x=y \rightarrow -(Ez)Ka(y=z)).
\]
And it may be easily demonstrated that (6), like (3), is perfectly consistent in KB. It looks, therefore, as though Hintikka has successfully defended KB against the charge that Restricted Range brings with it an inevitable restriction on the powers of expressibility.

Nevertheless this is not the end of the matter. Sleigh asks us to consider

(7) There is someone such that a knows that he exists but not who he is

in his paper 'Restricted Range in Epistemic Logic'. Making reference to the points Castaneda has raised he writes:

Still there may be life in this line of criticism. \[](3)\] and \[](6)\] are equivalent in KB. But the inference from \[](4)\] to \[](1)\] may seem counterintuitive. An examination of a proof that \[](3)\] implies \[](6)\] suggests that KB utilizes conditions which make it impossible to formulate an intuitively acceptable and consistent symbolization for \[](7)\]. It is natural to assume that \[](7)\] is consistent, that it is not correctly symbolizable in KB and that this failure is a consequence of the restricted range feature of KB.3

There is no disputing the fact that (3) and (6) are equivalent in KB; but why should this convince us that (7) cannot be consistently rendered in KB? Unless I am missing something obvious, it seems apparent that we can use Hintikka's ploy once again and find the consistent symbolization we are looking for. In the case at hand this would amount to

(8) \((Ex)((y)(x=y \rightarrow Ka(Ez)(z=y)) \& (y)(x=y \rightarrow -(Ez)Ka(z=y)))\).

Like both (3) and (6), (8) is perfectly consistent in KB. And so once again it appears that we have no English sentences whose failure to be expressed in KB can be pinned on the Restricted Range feature.
Castaneda's criticisms, though interesting, do not seem to amount to anything in the end. 4

Further criticism of the Restricted Range feature is raised by Romane Clark in his paper 'Comments'. His remarks arise out of a set of comments directed at Hintikka's semantics for perception, but the point he raises applies to KB in precisely analogous fashion.

Consider Existential Generalization in the following form:

\[
\text{Perceives } p(b/x) \in w \\
\text{(Ex)(b=x)} \in w \\
\text{(Ex)(Perceives } p(x) \in w. \\
\]

Again suppose that 'p' contains no epistemic operators.

It will have been noticed that Hintikka's account of the nature of the premise whose addition reconstitutes the validity of E.G. is different, and stronger, than that given above....Hintikka requires not merely that the object of the agent's sensuous belief exists, but (I think) that the agent perceives who he is \text{E(Ex)(Perceives b=x)}]. Are these stronger assumptions essential; i.e., do the weaker versions of E.G. sanction invalid inferences for the contexts of belief and perception? The answer is "no" on at least one transcription of the commonsense statements of belief or perception into a semantics close in spirit to Hintikka's own....Why then does Hintikka require the stronger premise...? It is not, I think, part of commonsense to do so....Hintikka, it seems to me, is led to his stronger requirement because of a stronger, and I believe doubt-ful, principle about how quantifiers are to be understood....This principle of "relative agent omniscience" seems to me to be false.5

Here Clark makes two important claims: (i) The employment of the weaker version of E.G., which for knowledge amounts to

\[
\text{Ka } p(b/x) \in w \\
\text{(Ex)(b=x)} \in w \\
\text{(Ex)Ka } p(x) \in w \\
\]

produces an adequate system, and (ii) The system produced avoids the principle of "relative agent omniscience" (as he refers to the Restricted Range feature), a principle which is "doubtful" and seemingly "false".
The first of these contentions we shall examine in the next section. We shall discover the system he proposes which lacks the principle of relative agent omniscience ("RAO" for short) is, while perhaps in some sense reasonable, quite a bit less interesting and powerful a system than KB. As far as his contention that the principle of RAO seems false, Clark has this to say:

Nonetheless, this principle seems to me to be pretty clearly unacceptable, for it precludes, or appears to preclude, our saying quite ordinary things. Take our freshman and Vercingetorix again. I want to say that there was someone, namely V., whom our freshman does not believe to have existed. But the natural symbolization of this, read back into English, comes out in Hintikka's translation to be that there was someone known to our freshman whom he believed not to exist.  

Here it is the belief version of Castaneda's (1) which, given the presence of the Restricted Range feature, seems to resist translation. The most natural way to symbolize

(9) There was someone our freshman does not believe to have existed,

namely,

(10) (Ex)¬Bf(Ey)(x=y),

turns out to be the symbolization for the longer sentence

(11) Someone known to our freshman is not believed by our freshman to have existed.  

Clark's criticism differs from Castaneda's in that, while the latter charged that the natural translation (2) of (1) is indefensible, Clark argues that what one might expect to be the natural translation of (9), viz., (10), is the translation already for an English sentence (11) not equivalent to (9). Nothing about consistency is mentioned in his criticism. Therefore, Hintikka's reply to Clark that (C.EX.=*) has been dropped from KB seems beside the point.
Whether or not one can prove that (10) is indefensible, the fact remains that (10) is a formal translation of (11) and hence cannot at the same time be a formal rendition of (9). What Hintikka should have pointed out is that the way to translate (9) is not the way that strikes one as "most natural", but rather,

\[(12) \ (Ex)(y)(x=y \rightarrow Bj(Ez)(y=z)).\]

And so, if the RAO principle is an undesirable feature of KB, it's undesirability does not lie, as far as Clark has shown, in precluding our saying "quite ordinary things". Restricted Range has simply not been shown to limit the expressibility of KB.

There is one further criticism of the Restricted Range feature which has appeared in print, but we shall note it only in passing. Sellars, in his 'Some Problems About Belief' levels the following charge against KB:

When Hintikka offers a definition...of a 'transparent' in terms of an 'opaque' sense of 'believe', the definition does not achieve the above purpose, for the range of the quantified variable is restricted to individuals known to the person whose beliefs are under consideration. Hintikka's claim to have defined "Quine's transparent sense" in terms of "the basic (opaque) sense plus quantification" is simply mistaken.\(^7\)

It is obvious that a satisfactory discussion of a charge such as this would require a lengthy digression into a number of profound and controversial areas relating to work done by Quine. Hence we shall not evaluate the charge, noting only that (i) Sleigh argues that Sellars's criticism relies upon attributing to Hintikka a definition of 'transparent knowledge' which is "not an accurate representation of Hintikka's account of transparency", and (ii) Quine points out that Sellars's own account of transparency leads to paradoxical results.\(^{10}\)
In this section we shall examine the highlights of several variations of KB which avoid the Restricted Range feature. This will include systems proposed by Fjellesdal in 'Knowledge, Identity, and Existence', by Tienson in 'The "Basic Restriction" in Hintikka's Quantified Epistemic Logic', and by Sleigh in 'Restricted Range in Epistemic Logic'.

As a preliminary to considering Fjellesdal's system we shall talk briefly of Clark's claim, mentioned above, that an adequate revision of a Hintikka-type semantics can be produced by employing Existential Generalization in its "weak form", which for knowledge is:

\[
\begin{align*}
\text{Ka } p(b/x) & \in w \\
(\text{Ex})(b=x) & \in w \\
(\text{Ex})\text{Ka } p(x) & \in w.
\end{align*}
\]

Suppose we revise KB in such a way that (108) and (109) are replaced by:

\[
\begin{align*}
\text{(108C)} & \quad (\text{Ex})\text{Ka } p(x) \in w \\
(\text{Ex})(b=x) & \in w \\
\text{Ka } p(b/x) & \in w \\
\text{(109C)} & \quad (x)\text{Ka } p(x) \in w \\
(\text{Ex})(b=x) & \in w \\
\text{Ka } p(b/x) & \in w.
\end{align*}
\]

Clearly such a system will yield Existential Generalization in its weak form. Is such a revision of KB promising?

At first glance it appears that exactly the opposite is the case. Consider Existential Generalization. If Jones knows that the tallest spy is a spy, and there is in fact a tallest spy, is it plausible to conclude that the tallest spy is known by Jones to be a spy (and hence that someone is known by Jones to be a spy)? Can we conclude that Jones knows with respect to someone that he's a spy?

It's obvious that Clark does not intend to sanction such inferences. Suppose we are able to draw a distinction between what
people have called "de re" and "de dicto" senses of knowing; then we might say that Clark's purpose is not to propose that de dicto knowledge together with an existence presupposition yields de re knowledge. Then how might Clark avoid the inference in question? One method would be to limit the constant terms in the language in such a way that 'the tallest spy' does not automatically qualify as a genuine constant term upon which one can generalize and instantiate. And another method would be to assign English readings to formulas in such a way that

\[ \text{Ka(}\text{the tallest spy is a spy}) \]
\[ (\text{Ex})(x \text{ is the tallest spy}) \]
\[ (\text{Ex})\text{Ka(x is a spy)} \]

comes out plausible. Of these, Clark chooses the second.

Consider '(Ex)Ka(x is a spy)'; by reading it "There exists someone who is known by a to be a spy" the inference is quite implausible. But I believe there is a reading of '(Ex)Ka(x is a spy)' according to which the inference is perfectly plausible: "There exists someone x and a knows that the proposition that x is a spy is true". We do not require for the truth of '(Ex)Ka(x is a spy)' that anyone be known by a, according to the latter reading. In other words, we simply do not give '(Ex)Ka(x is a spy)' a de re reading; rather, we read it de dicto. As unusual a practice as this might be, it does, I believe, give us a way to make perfectly good sense out of Clark's proposal. By giving the conclusion of the above inference a de dicto reading, we move only from a de dicto premise to a de dicto conclusion, and the inference is perfectly plausible. Given this way of reading '(Ex)KaFx' and '(x)KaFx' there appears nothing wrong with the adoption of (108C) and (109C); the rules themselves seem perfectly plausible.

Why then cannot one bypass the question of Restricted Range altogether by assigning de dicto readings to formulas formed
by quantifying in? Of course one can if all that's desired is a semantics for *de dicto* knowledge. The system Clark seems to be proposing is, as far as I can see, not of much interest at all. By leaving no room for the *de re* sense of knowing one is perhaps able to produce a system that is in some sense perfectly acceptable. But such a system appears to be only mildly interesting and certainly a much less powerful system than KB.

Suppose we take the opposite course and revise KB in such a way that (1) (108) and (109) are replaced by (108C) and (109C), and (ii) To avoid implausible inferences we place a limitation on terms upon which one is allowed to instantiate and generalize in such a way that 'the tallest spy' does not automatically qualify. Just such a proposal forms the basis of a system proposed by Føllesdal.

Consider the following well known difficulty.

(a) The man who is coming toward me = Corsicus
(b) I know that the man who is coming toward me is musical
(c) I know that Corsicus is musical

Although the principle of substitutivity of identity in epistemic contexts licenses the inference from (a) and (b) to (c), the inference is obviously implausible. In this connection Føllesdal observes:

> It seems...that we are in a dilemma, on the one side we have the *prima facie* implausibility of the principle of substitutivity of identity in epistemic contexts, on the other side we have the metaphysical and epistemological reasons for assuming the principle. Hintikka, in his book, chooses to reject the principle. In this paper, I shall choose the other alternative and accept the principle. 11

To explain how the acceptance of the principle is compatible with the
failure of (a) and (b) to imply (c), Føllesdal asks us to consider

(13) \((\forall y)K_a(x)(Fx \iff (x=y))\).

The fact that the expressions 'the man who comes towards me' and 'Corsicu' change their reference from world to world...should perhaps be taken as evidence that they contain some descriptive element, and that they shouldn't be regarded as genuine names. The only descriptions that should be regarded as genuine names are those which keep the same descriptum in every possible world, that is, in symbols, a description \((\forall x)Fx\) behaves like a genuine name if and only if \(\mathcal{L}(13)\). It may be shown that we can avoid all the difficulties that we have been discussing till now, if we require all names to satisfy this condition, i.e., permit a name-like expression \(a\) to be treated like a name only if it satisfies \((\forall y)K_a(x)(x a \iff (x=y))\). Only expressions which satisfy this condition should be permitted to flank identity signs and be used in arguments turning on existential generalization and universal instantiation. All other name-like expressions should be regarded as hidden descriptions and be eliminated, e.g., by Russell's theory of definite descriptions. 12

We have, therefore, a system which (unlike KB) preserves substitutivity of identity but avoids bad inferences by a restriction on constant terms which may occur in well-formed formulas. Let 'b' be short for 'the tallest spy'. Then \((\forall y)K_a(x)(x b \iff (x=y))\), if true, enables us to symbolize "a knows that the tallest spy is a spy" as 'K_a(b is a spy)'. On this symbolization the inference goes through. On the other hand, if \((\forall y)K_a(x)(x b \iff (x=y))\) is false, then the first premise is symbolized 'K_a(\(\forall y\))(x (x b \iff (x=y)) & y is a spy)', a formula which contains no singular terms other than variables, and with this symbolization the inference is blocked.

Similar remarks apply to the Corsicus example.

Does Føllesdal's system (let's call it 'F') lack the Restricted Range feature? It is absolutely clear that Føllesdal intends F to avoid Restricted Range. On page 14 he indicates that
'(x)Ka(Ey)(y=x)' is to be read "Every object is such that everybody knows that it exists". And on page 23 he points out that '(x)KaFx', which is for Hintikka read, "Of each person which a knows, a knows that it is F", is read in F as, "Each person is known by a to be F". Thus '(x)KaFx' and '(x)(Ka(x=x) Æ KaFx)' will not be equivalent in F.

Given that Füllesdal intends F to lack the Restricted Range feature, the question is whether this can be shown to be the case in F. The answer is "yes", provided certain natural assumptions are made about F. These assumptions can be made explicit in what we might call the "large-scope correlates" of (108C) and (109C):

(108F)  
(Ex)KaFx Æ w
(Ex)(y)(x b's Æ (x=y)) Æ w
(Ex)(y)((x b's Æ (x=y)) & KaFx) Æ w, for some predicate 'b's'

(109F)  
(x)KaFx Æ w
(Ex)(y)(x b's Æ (x=y)) Æ w
(Ex)(y)((x b's Æ (x=y)) & KaFx) Æ w.

It is absolutely essential that F contain these rules; for without them F is incapable of accounting for such inferences as:

Everyone is known by a to be F
The tallest spy is someone
The tallest spy is known by a to be F,

when a fails to know who the tallest spy is.

But given these rules it is easy to show that F lacks the Restricted Range feature. To show this it is sufficient to show that

(x)KaFx Æ w
(x)Fx Æ w

is valid. Assume both that '(x)KaFx' Æ w and that '(Ex)-Fx' Æ w. By E.I., '-F(7 x)(x b's) Æ w. Now if '(Ey)Ka(x)(x b's Æ (x=y)) Æ w, then '-Fb' Æ w, and hence '(Ex)(b=x) Æ w. So by U.I., 'KaFb Æ w, from which a contradiction immediately follows. On the other hand, if '(Ey)Ka
(x)(x b's\leftrightarrow(x=y))' \in w, then \'(Ex)(y)((x b's\leftrightarrow(x=y)) \& -Fx)' \in w. Hence 
'(Ex)(y)(x b's\leftrightarrow(x=y)) \in w; but by (109F) ')(Ex)(y)((x b's\leftrightarrow(x=y)) \& KaFx)' \in w. And it can easily be shown that ')(Ex)(y)((x b's\leftrightarrow(x=y)) \& -Fx)' and ')(Ex)(y)((x b's\leftrightarrow(x=y)) \& KaFx)' entail a contradiction. So in either case the assumption we began with entails a contradiction.

There is one rather surprising feature found in F. Although 'KaFx\rightarrow-Pa-Fx' is valid in F, '\-'Pa-Fx\rightarrow KaFx' is not.\(^{13}\) According to Føllesdal, 'KaFx' is true just in case 'x is F' is true in every compatible world; but in order for 'x is F' to have a truth-value in a world, x must exist in that world. If, therefore, there are some compatible worlds in which x fails to exist, 'KaFx' is false. Yet it may still be true that in no compatible worlds is 'x is F' false, and if so, '\-'Pa-Fx' will be true. Hence '\-'Pa-Fx' and '\-'KaFx' may be embedded in the same model set.

Føllesdal's reasons for formally distinguishing 'KaFx' and '\-'Pa-Fx' in F arise out of the following difficulty. It would appear that unrestricted substitutivity of identity would make formulas of the form 'b=c\rightarrow Ka(b=c)' valid. But by simple quantification theory one can infer from 'b=c\rightarrow Ka(b=c)' both ')(x)(y)(x=y\rightarrow Ka(x=y))' and ')(x)Ka(Ey)(x=y)'.\(^{14}\) And in a system lacking Restricted Range these formulas have counterintuitive readings (more counterintuitive than the already counterintuitive readings with Restricted Range). Therefore to avoid the validity of ')(x)(y)(x=y\rightarrow Ka(x=y))' and ')(x)Ka(Ey)(x=y)', F will require substitutivity of identity only insofar as it requires the validity of the weaker condition 'b=c\rightarrow Pa-(b=c)'; in compatible worlds in which b and c both exist 'b=c' will be true, and otherwise 'b=c' will lack a truth-value. Now we need only accept the validity of the weaker formulas ')(x)(y)(x=y\rightarrow Pa-(x=y))' and ')(x)Pa-(Ey)(x=y)', formulas which Føllesdal believes to have
perfectly innocent readings.

The basic idea behind the solution is that when we have now eliminated all spurious singular terms and sit back with merely variables, then whenever these variables, which function approximately like the 'it' of ordinary language, have no reference, sentences which contain them have no truth value.\(^5\)

By admitting truth-value gaps and by not requiring the validity of modal negation, therefore, Føllesdal believes he has produced an adequate system of epistemic logic which, as it turns out, lacks the Restricted Range feature. We return to Føllesdal's system in the next section.

A rather interesting modification of Føllesdal's F lies at the heart of a system proposed by J. Tienson in 'The "Basic Restriction" in Hintikka's Quantified Epistemic Logic'. In this paper Tienson sets out to construct a system of epistemic logic using model sets which "preserves Hintikka's insights concerning quantification into epistemic contexts but which does not contain the basic restriction [i.e., the Restricted Range feature]".\(^16\)

Considered as a revision of Hintikka's KB, Tienson's system -- let's call it $F^*$ -- is succinctly described as follows:

What is required is two different kinds of individual symbols...Let us allow ordinary constants to continue to represent ordinary singular terms. Thus we retain \(L(109)\) and drop \(L(108)\). Let us then introduce a new type of symbol, say auxiliary symbols, which can be thought of as purely designative. All the rules will apply to formulas containing auxiliary symbols -- auxiliary formulas -- but they need not be thought of as having meaning in themselves. Three new rules dealing with auxiliary symbols are needed.

- **(C.E.\(_A\))** If \((\exists x)A^M\), then \(A(\alpha/x)^M\) and \((\exists x)(x=\alpha)^M\), for at least one auxiliary symbol \(\alpha\).
- **(C.U.\(_A\))** If \((x)A^M\) and \((\exists x)(x=\alpha)^M\), then \(A(\alpha/x)^M\).
- **(C.-\(_A\))** If \(\alpha = \beta^M\) and \(\beta^*\) is alternative to \(\alpha\), then \(\alpha = \beta^M^*\). 


In addition, Hintikka's rules of substitutivity of identity in atomic formulas (C.=.), must be extended to refer to auxiliary symbols as well as constants. The case not yet covered is that of identity statements involving one constant and one auxiliary symbol. When these rules are added to Hintikka's rules for propositional epistemic logic the resulting system does not contain any restriction of values of variables, but it does preserve Hintikka's insights concerning the relationship between constants and quantified formulas.17

It appears to me that F* and F may be contrasted in the following way. Where auxiliary symbols are employed in F*, genuine names are employed in F; and where ordinary constants are employed in F*, expanded descriptions are employed in F. Furthermore, F* does not contain truth-value gaps, and F* preserves the validity of modal negation. But in other respects, however, F* and F do not seem to differ at all. Unqualified substitutivity of identity is preserved in F* (relative to auxiliary symbols), the Restricted Range feature is absent, and existential generalization in Clark's "weak form" is valid.

Tienson's system, by eliminating truth-value gaps and preserving the validity of \( \neg P a \neg F x \ Ka F x \) does seem to be a more natural system than Føllesdal's F. But one must recall that Føllesdal forfeits modal negation for the sole purpose of avoiding certain "bad theorems" which one encounters as a result of his basic approach. Since Tienson's basic approach does not differ significantly from Føllesdal's, and since no attempt is made by Tienson to avoid bad theorems, it would be reasonable to suppose that these theorems show up in F*.

As we shall see in the next section, this is indeed the case.

The final variant of KB minus Restricted Range we shall consider is Sleigh's system K which he sets forth in 'Restricted Range in Epistemic Logic' as well as 'Epistemic Logic, Again'. Sleigh
sets out his system in the form of a model theory, but it can conveniently be compared to the theory of model sets of KB in a manner we shall explain. The model theory itself (K) is very elegantly described in the following way:

**Elements of a model**

W - a non-empty set of worlds

D - a non-empty set of individuals

R - a reflexive and transitive relation defined over W

Q - a function assigning to each world a subset of D so that

\[ Q(w_i) = D \]

I and V - interpretation and valuation functions characterized below

**Interpretation functions** must satisfy these conditions:

(i) For each variable x and each world \( w_i \), \( I(x, w_i) = u \) for some \( u \) in D with the proviso that \( I(x, w_i) = I(x, w_j) \) for all \( w_i \) and \( w_j \).

(ii) For each non-variable term \( t \) and each world \( w_i \), \( I(t, w_i) = u \) for some \( u \) in D. Here we allow for the possibility that \( I(t, w_i) \neq I(t, w_j) \).

(iii) For each n-ary predicate \( \phi \) and each \( w_i \), \( I(\phi, w_i) \) is a set of n-tuples from D.

Each interpretation function determines a unique **valuation** function satisfying these conditions:

(i) **Identities:** For any terms \( \alpha \) and \( \beta \) and any world \( w_i \),

\[ V_I(\alpha = \beta, w_i) = 1 \text{ iff } I(\alpha, w_i) = I(\beta, w_i). \]

(ii) **Atomic formulas:** Where \( \phi \) is an n-ary predicate and \( \alpha_1, \ldots, \alpha_n \) are terms \( V_I(\phi, \alpha_1, \ldots, \alpha_n, w_i) \) is iff

\[ \langle I(\alpha_1, w_i), \ldots, I(\alpha_n, w_i) \rangle \in I(\phi, w_i). \]
Negations, implications, and epistemic statements are treated as one would expect -- e.g.,

(iii) (a) \( V_I(Ka\mathcal{U},w_1) = 1 \) iff for every \( w_j \) such that \( w_1 R w_j \),
\[
V_I(\mathcal{U},w_j) = 1.
\]
(b) \( V_I(Pa\mathcal{U},w_1) = 1 \) iff for some \( w_j \) such that \( w_1 R w_j \),
\[
V_I(\mathcal{U},w_j) = 1.
\]

(iv) \( V_I((x)\mathcal{U},w_1) = 1 \) iff for any \( I' \) if
(a) \( I' \) is like \( I \) except (possibly) at \( x \), and
(b) \( I'(x,w_1) \in D_i \), then
\[
V_I(\mathcal{U},w_1) = 1.
\]

We may note then any tuple \( \langle W,D,R,Q,I \rangle \) meeting these conditions is a model for our simplified epistemic logic, that \( V_I(\mathcal{U},w_1) = 0 \) in case \( V_I(\mathcal{U},w_1) \neq 1 \) and that \( \mathcal{U} \) is valid just in case for each \( I \), \( V_I(\mathcal{U},w_1) = 1 \) for each world \( w_1 \) in each model.\(^{18}\)

It is very easy to describe how \( K \) differs from \( KB \). Form the system \( H \) by the following modification of \( K \): When (and only when) the variable \( x \) occurs within the scope of an epistemic operator and is bound from the outside by \( '(x)' \), require not only that \( I'(x,w_1) \in D_i \) but that \( I'(x,w_j) \in D_j \), for all \( w_j \) such that \( R w_1 w_j \). In all other cases proceed as in \( K \). According to Sleigh, "The model theoretic system \( H \) has the same theorems as the model set system \( KB \)."\(^{19}\)

It is easy to show that \( K \) lacks the Restricted Range feature. We show that the move from \( '(x)KaFx' \) to \( '(x)Fx' \) is valid in \( K \). Let \( w_1 \) be any member of \( W \). If \( V_I((x)KaFx,w_1) = 1 \), then for any \( I' \) just like \( I \) except (possibly) at \( x \), if \( I'(x,w_1) \in D_i \), then \( V_I(KaFx,w_1) = 1 \).

Therefore, \( V_I(Fx,w_1) = 1 \), for all \( w_j \) such that \( R w_1 w_j \). But since \( R w_1 w_1 \) holds by the reflexivity of \( R \), \( V_I(Fx,w_1) = 1 \). Therefore, for any \( I' \)
just like I except (possibly) at x, if I'(x, w_i) \in D_i, then V_{I'}(Fx, w_i) = 1. But this holds just in case V_I((x)Fx, w_i) = 1, which was to be proved. Notice that a similar strategy fails in H. It can be shown that for any I' just like I except (possibly) at x, if I'(x, w_j) \in D_j, for all w_j such that R_{w_i}w_j, then V_{I'}(Fx, w_j) = 1. But from this it does not follow that V_I((x)Fx, w_i) = 1. For some I* just like I except (possibly) at x may map (x, w_i) to a member of D_i, fail to map (x, w_j) to a member of D_j, where R_{w_i}w_j, and be such that V_{I*}(Fx, w_i) = 0. It will then be false that for every I' just like I except (possibly) at x, if I'(x, w_i) \in D_i, then V_{I'}(Fx, w_i) = 1.

Having taken a brief look at the systems proposed by Føllesdal, Tienson, and Sleigh, we are now ready to say a few words about each.

Turning first to Føllesdal's system F, we may note that there certainly are no obvious difficulties to be found. His basic approach, preserving unqualified substitutivity of identity by eliminating from well-formed formulas any singular term 'b' which fails to satisfy '(\exists y)(x)(x b's<->(x=y))', is no doubt regarded by many logicians as the most preferred approach to handling singular terms in quantified epistemic logic. Furthermore, his proposal to allow truth-value gaps together with his proposal to invalidate '-Pa-Fx\rightarrow KaFx' in F is, on the surface of things, not the cause of any real difficulty. Finally, what one might expect to be theorems in F, '(x)(y)(x=y\rightarrow Ka(x=y))' and '(x)Ka(Ey)(x=y)', formulas with counter-intuitive English readings, are not theorems at all.

In spite of all of this, Stine feels that all is not well with Føllesdal's F. Consider:

(a) (x)(y)(x=y\rightarrow Pa-(x=y))
(b) \((x)\neg Pa-(Ey)(x=y)\)

(c) For all \(x\) and all \(y\), if \(x\) and \(y\) are identical, then it is not compatible with \(a\)'s knowledge that they are not identical.

(d) Everything is such that it is not compatible with \(a\)'s knowledge that it does not exist.

Given readings (c) and (d) of (a) and (b), respectively, it is well that we do not require (a) and (b). Føllesdal tries to justify (a) and (b) with the uniform reading of quantifiers, but his argument is tortured. To make his point he must give up the convertibility of "\(Ka\)" and "\(\neg Pa\)" and hold that "\(a\) exists" is not false, but without truth-value, if \(a\) does not exist. The latter seems clearly objectionable; it is one thing to deny that "\(Fa\)" has a truth-value if \(a\) does not exist, but another thing to deny that it is false that \(a\) exists.  

I believe that Stine's remarks point us in the direction towards uncovering two features of \(F\) which do seem to be difficulties.

Stine's first suggestion is that (even with the invalidity of 

\[(a') (x)(y)(x=y \rightarrow Ka(x=y)), \text{ and} \]

\[(b') (x)Ka(Ey)(x=y)\]

in \(F\) formulas (a) and (b), which are valid in \(F\), have English readings which are themselves counterintuitive. This point seems to me to be perfectly correct. Sentences (c) and (d) just are equivalent versions of the sentences

(e) For all \(x\) and all \(y\), if \(x\) and \(y\) are identical, then \(a\) knows \(x\) and \(y\) are identical, and

(f) Everything is such that \(a\) knows that it exists, respectively. So there is no reason in the claim that (c) and (d) are not counterintuitive even though (e) and (f) are. Why then does Føllesdal think that (a) and (b), unlike their knowledge counterparts (\(a'\)) and (\(b'\)), have readings that are perfectly innocent?

The reason is quite simple. The readings Føllesdal assigns
to (a) and (b) are (roughly):

(g) For all x and all y, if x and y are identical in the actual world, then in no compatible world is 'x=y' false

(h) For any x in the actual world, there is no compatible world in which 'x exists' is not true.

It is true that (g) and (h) are perfectly innocent, for (g) and (h) are descriptions (in English) of the truth-conditions of (a) and (b) in F, and these truth-conditions make perfectly good sense. But what is the relevance of (g) and (h) to the question whether (a) and (b), like (a') and (b'), have English readings which are counterintuitive? None as far as I can see. One must distinguish between the English reading of a formula and a description (in relatively casual English) of a formula's truth-conditions; just because (g) and (h) are not counterintuitive, one cannot conclude that (a) and (b) are no trouble to admit as theorems. It seems undeniable that (a) and (b) are no better to admit as theorems than (a') and (b'). Hence it looks as though F has "bad theorems" after all.

Stine goes on to suggest that problems arise in connection with sentences of the form "b does not exist". Let us identify an example of such a problem. Consider:

(14) a knows that b does not exist

(15) Ka-(Ex)(b=x)

(16) So far as a knows, b does not exist

(17) Pa-(Ex)(b=x).

It is a peculiar feature of F that (15) and (17) are inconsistent. Formally speaking, of course, this feature is unproblematic, but one is tempted to suppose that the inconsistency of (15) and (17) may raise problems in the expressibility of English sentences within
the resources of F. In particular, one might suspect that (14) and (16) cannot be consistently rendered in F.

It may be replied that, although (15) and (17) are indeed inconsistent, the following are not:

(18) \(-Pa(Ex)(b=x)\)

(19) \(-Ka(Ex)(b=x)\).

Hence, it may be urged, one may symbolize (14) as (18) and (16) as (19). It is true that (18) is consistent, and thus one may regard (18) as a perfectly adequate symbolization of (14). But (19), on the other hand, is not consistent in F. Therefore, one cannot consistently render (16) as either (17) or (19); thus it looks as though F is incapable of handling sentences such as (16), and if this is true it's a serious difficulty of F.

Tienson's system F* avoids both of these difficulties for the very simple reason that the validity of modal negation is preserved and truth-value gaps are not introduced. It would seem, therefore, that the problems which drove Føllesdal to the somewhat radical proposals he made will not be avoided by Tienson. And this turns out indeed to be the case; the following are all theorems in F*:

\[
\begin{align*}
(x)(Ey)Ka(x=y) \\
(x)(y)(x=y \rightarrow Ka(x=y)) \\
(x)Ka((Ey)(x=y)) \\
(x)Ka(x=x) \\
(x)Ka(Fx \lor -Fx)
\end{align*}
\]

And each of these has an English reading which is quite counterintuitive. Consider, for example, '\((x)Ka(x=x)\)'. Suppose that for some model set M, '\(-(x)Ka(x=x)\) \in M'. Then '\((Ex)Pa(x/x)\) \in M'; so by (C.E.\(_A\)) there is an auxiliary symbol '\(\alpha\)' such that '\(Pa(\alpha \neq \alpha)\) \in M'. Therefore, in some alternative M* of M, '\(\alpha \neq \alpha\) \in M*'. But
this is impossible, and hence our initial assumption is false.
Proofs of other formulas proceed along similar lines.

Sleigh's system too makes theorems out of the formulas listed above (with the exception of '(x)Ka(Ey)(x=y)'). Consider, for example, '(x)Ka(x=x)'. If $V_I((\exists x)Pa(x \neq x), w_1) = 1$, for some function $I$ and world $w_1 \in W$, then for some $I'$ just like $I$ except (possibly) at $x$, $I'(x,w_1) \in D_i$ and $V_{I'}(Pa(x \neq x), w_1) = 1$. So $V_{I'}(x \neq x, w_j) = 1$, for some $w_j$ such that $Rw_1w_j$. Hence $I'(x,w_j) \neq I'(x,w_j)$, but this is impossible by the definition of an interpretation function in $K$. Proofs of the other formulas proceed along similar lines.

Both $F^*$ and $K$ strike me as very reasonable attempts to avoid the Restricted Range feature. But it would seem worthwhile to investigate whether one could not pursue an alternative approach to avoiding Restricted Range which did not involve the bad theorems one finds in $F^*$ and $K$. In the next chapter I shall undertake such an investigation and conclude that one can design a model set theoretic system based upon $KB$ which avoids Restricted Range but at the same time avoids the unpleasant consequence that a has de re knowledge regarding everyone who exists.
APPENDIX

The general question we have been pursuing in this chapter is the question whether one can take an approach to epistemic logic similar to that taken by Hintikka and design a system lacking the Restricted Range feature. Føllesdal, Tienson, and Sleigh appear each to have shown that this can be done. There is, however, a system of epistemic logic proposed by Gail Stine in 'Quantified Logic for Knowledge Statements' which, though it lacks the Restricted Range feature, may not be fairly called a "variant" of Hintikka's KB. For this reason we shall consider it separately.

Stine's idea is to propose an "alternative approach to epistemic logic" which lacks features "of at least debatable desirability" found in KB, Føllesdal's F, and Sleigh's K. These features are three in number: Restricted Range, Deductive Closure, and Knowing a Name (the feature a system has just in case an agent a knows who an individual is only if, for some name 'b', a knows who b is). According to Stine, KB possesses all three of these features, K has the latter two, and F has the second. Stine's system -- let's call it 'S' -- is designed to avoid all three.22

It may raise some eyebrows to hear that Stine proposes to avoid Deductive Closure, but the method she employs is relatively unspectacular. Unlike all of the systems we have considered, S is simply to be thought of as a formal theory devoid of any intended interpretation. In particular, S is not to be thought of as having an intended "possible worlds interpretation".23 By divorcing S from possible worlds in this way, it can easily be shown, S removes from itself the stigma of
Deductive Closure. But clearly S is in this way a vastly different kind of system from all of the others we have up until now considered.

Because one is not to think of modal formulas in S in terms of possible worlds, it is almost impossible to develop more than the most elementary of intuitions regarding the nature of S. Moreover, in no place does Stine have much to say about the rules and axioms of S. We are told that \( (x)(x = x) \), \( (x)(y)(x = y \rightarrow (Fx \rightarrow Fy)) \), and \( (x) (KaFx \rightarrow Ka(x=x)) \) are axiom schemata. And rules for instantiation and generalization are described (U.I. and E.I. are exactly as in KB except that in E.I. a free variable is used in place of a new constant to avoid the "Knowing a Name" feature described above). But other than just this, everything else remains pretty much unspecified.

A good deal of Stine's concern in 'Quantified Logic for Knowledge Statements' is directed towards a critique of Føllesdal's F and Sleigh's K. We have already made reference to some insightful remarks Stine makes along these lines. It is, however, regrettable that not more explanation is provided relative to her own system S.

It may seem irrelevant to our purposes to say anything more about S. Nevertheless, Stine does include in her paper a list of English translations for key formulas in S, and a few words about these will, I believe, prove instructive towards an understanding of the system we shall examine in the next chapter. The translations go as follows:

(a) \( (x)KaFx \)  
Everyone is known by \( a \) to be \( F \), and
Everyone is known to \( a \) and known by \( a \) to be \( F \)

(b) \( (x)(Ka(x=x) \rightarrow KaFx) \)  
Everyone known to \( a \) is known by \( a \) to be \( F \)
Someone is known by \( a \) to be \( F \), and
Someone is known to \( a \) and known by \( a \) to be \( F \)

(c) \( (Ex)KaFx \)  
Everyone is known by \( a \) to be \( F \)
The English readings Stine provides seem to make good sense for a system which lacks Restricted Range. However, there are some apparent difficulties which deserve a word of explanation. It may appear strange that the formula '[(x)KaFx]' may be read either as "Everyone is known by a to be F" and "Everyone is known to a and known by a to be F". However, I believe this is as it should be, for it seems quite clear that these two English sentences are equivalent. Along the same lines, it turns out that formulas (f) and (g) are equivalent in S, and it may seem objectionable that the former is read "No one is known by a to be F" while the latter is read "No one known to a is known by a to be F". It's not obvious that these sentences are equivalent, but to argue that they are non-equivalent one must maintain that "Someone not known to a is known by a to be F" can be true (the reading for the negation of Stine's axiom '[(x) (KaFx→Ka(x=x)])', and it's hard to see how that line might be defended. Finally, it might be argued, if "Everyone is known to a and known by a to be F" is an admissible reading of (a), then by quantifier negation it would seem that (d) ought to be read not as "Someone is not known by a to be F" but rather "Someone is either not known by a by a to be F or not known to a. And once again Stine seems to be in the clear. These latter two sentences are perfectly equivalent once we accept the truth of "Everyone known by a to be F is known to a", as I believe we must.
All in all, Stine's readings seem perfectly correct and acceptable, despite initial appearances. Hence, quite apart from anything I have to say about Stine's system S, it seems to me that Stine's lexicon is worth remembering. I believe that this lexicon, or something closely resembling it, is quite inevitable for a system which lacks the Restricted Range feature. We shall see more of these readings in the next chapter.
FOOTNOTES

1Hintikka, 'Individuals, Possible Worlds, and Epistemic Logic', p. 34.

2Ibid., p. 37.

3Sleigh, 'Restricted Range in Epistemic Logic', p. 69.

4Sleigh has pointed out in conversation that, even without (C.EK.=*),
it can still be proven that (6) virtually implies (3). And since
(4) certainly does not imply (1), it looks as though there perhaps
are still some counterintuitive results which can reasonably be
blamed on Restricted Range. This point seems to me perfectly correct.

5Clark, 'Comments', pp. 177-179.

6Ibid., p. 179.

7Clark seems to be mistaken about requiring that someone be known
(as opposed to someone's identity being believed to be something
or other), but nothing seems to hang on this oversight; also,
nothing seems to hang on tenses in this example.


9Sellars, 'Some Problems About Belief', p. 191.

10Sleigh, 'Restricted Range in Epistemic Logic', p. 74; Quine, 'Reply
to Sellars', p. 377.


12Ibid., pp. 11-12.

13Apparently open formulas are well-formed in F (unlike KB).


16Tienson, 'The "Basic Restriction" in Hintikka's Quantified Epistemic
Logic', p. 1.

17Ibid., pp. 7-8.


19Ibid., note 6.

20Stine, 'Quantified Logic for Knowledge Statements', p. 137.
21 The formula '(x)Ka(Ey)(x=y)' is a theorem only given a rule which Tienson is not sure whether to employ (see p. 8).


23 Ibid., p. 130.
CHAPTER V

In this chapter I shall propose a modification of Hintikka's system KB which lacks the Restricted Range feature. We have already looked at variations of KB proposed by Føllesdal, Tienson, and Sleigh, each of which lacks the Restricted Range feature, but we have found that each of these systems is beset by a number of undesirable theorems. In the system I propose all of these theorems are avoided by placing certain constraints upon the instantiation of existentially quantified formulas, constraints which we shall argue to be neither arbitrary nor unnatural.

We shall conclude that Restricted Range can be avoided in a Hintikka-type semantics without creating any new difficulties not present in KB. In closing, we return to Hintikka's argument (presented in the last chapter) that Restricted Range is an inevitable consequence of his approach to epistemic logic, and we shall find that the proof rests upon a faulty assumption regarding the role of instantiation and generalization in epistemic logic.

* * *

We recall that the following formulas are theorems in both Tienson's system F* and Sleigh's system K:

(a) \( (x)Ka(x=x) \)

(b) \( (x)(Ey)Ka(x=y) \)

(c) \( (x)(y)(x=y \rightarrow Ka(x=y)) \)

(d) \( (x)Ka(Fx \lor -Fx) \).

It can easily be shown that all of these formulas are provable in KB.\(^1\)
But there is a significant difference in the way these formulas are read in KB, on the one hand, and in F* and K on the other. In KB these formulas are read, respectively, as

(1) Everyone known to a is known by a to be self-identical
(2) For everyone known to a there is someone known to a such that they are known by a to be identical
(3) Any two persons known to a are, if identical, known by a to be identical
(4) Everyone known to a is known by a to be either F or not-F.

It may be a matter of dispute whether these readings are counterintuitive. Elsewhere I have argued that (3) is counterintuitive, but it's not at all clear that the same can be said for the others; (2), for example, seems on the surface to be entirely plausible.

The story is much different, on the other hand, when we contrast (1)-(4) with

(1') Everyone is known by a to be self-identical
(2') For everyone x there is someone with whom a knows x is identical
(3') Any two persons are, if identical, known by a to be identical
(4') Everyone is known by a to be either F or not-F.

These are, respectively, the readings of (a)-(d) in F* and K. It's absolutely obvious that (1')-(4') are each entirely implausible. While (3) is somewhat objectionable, its implausibility is mild in comparison to (3'), which, according to Føllesdal, says that a "...knows the right answers to all questions of identity". And it is a consequence of (1'), (2'), and (4') that a has de re knowledge regarding everyone who exists, not just those known to a.

It is, of course, the presence or absence of the Restricted
Range feature which determines whether (a)-(d) take on the troublesome readings (1')-(4') or the much less troublesome readings (1)-(4). Since KB has Restricted Range the world-lines over which '(x)' ranges do not pick up every individual in the domain of the actual world; they pick up only that subset of individuals which Hintikka describes as those who are "known" to a. Loosely speaking, '(x)' might be described as ranging over that subset of individuals who are known to a. In F, F*, and K, on the other hand, the world-lines over which '(x)' ranges do succeed in picking up every individual in the domain of the actual world. And in the same loose manner of speaking, '(x)' might be said to range wholesale over the individuals who comprise the domain of the actual world.

This phenomenon constitutes a good case for Hintikka's arguing that Restricted Range ought to be present in a Hintikka-type semantics for knowledge (whether or not it must be present on independent logical grounds). If the theoremhood of (a)-(d) is unavoidable in a Hintikka-type semantics, then to avoid the highly counterintuitive readings (1')-(4') of (a)-(d) it is greatly to one's advantage (other things being equal) to have Restricted Range.

Much the same goes for Föllesdal's F. Although none of (a)-(d) are provable in F, the following are:

(a') (x)-Pa-(x=x)
(b') (x)(Ey)-Pa-(x=y)
(c') (x)(y)(x=y → Pa-(x=y))
(d') (x)-Pa-(Fx v -Fx).

And while (a')-(d') are likewise provable in KB and have readings which are relatively innocent in KB, the readings are not so innocent in F.
Again, the innocence of (a')-(d') in KB is solely attributable to the presence of the Restricted Range feature. Thus it is open for Hintikka to argue that, given the theoremhood of (a')-(d'), it is to one's advantage not to do away with the Restricted Range feature.

It looks, therefore, like there are good reasons not to propose variations of KB which lack Restricted Range as long as either (a)-(d) or (a')-(d') are theorems. 4

***

Can a plausible system be produced after the fashion of KB which not only lacks Restricted Range but which fails to make theorems out of (a)-(d) and (a')-(d')? Let us begin by investigating just why systems based upon KB are so prone to make theorems out of (a)-(d) and (a')-(d') in the first place. Suppose we concentrate for the time being upon (a) and see how it's proved in KB:

(a) \( -(x)Ka(x=x) \equiv w \) Redog

(b) \( (Ex)Pa(x \neq x) \equiv w \) (a),(C.-U),(C.-K)

(c) \( Pa(b \neq b) \equiv w' \) (b),(108), for some 'b'

(d) \( (Ex)Ka(b=x) \equiv w' \) (c),(C.P*), for some w* such that \( R_{w*} \)

(e) \( b \neq b \equiv w* \) (C.self.=)

(f) \( b=b \equiv w* \)

The crucial move in this proof is clearly the inference from step (b) to (c) and (d) by way of rule (108). Let us closely examine the rule which allows this inference. Hintikka's rule (108) is usually thought of as a license which allows one to move from formulas of the form \( (Ex)KaFx' \) to \( KaFb' \) and \( (Ex)Ka(b=x)' \), for some 'b', in some given model set w. But it is at the same time a license which
allows one to instantiate upon formulas of the form \( (Ex)PaFx \):

\[
\begin{align*}
(Ex)PaFx & \in w \\
PaFb & \in w \\
(Ex)Ka(b=x) & \in w, \text{ for some } 'b'.
\end{align*}
\]

This latter consequence of rule (108) has been the object of criticism. Tienson, for example, writes:

But \( E(108) \) also says that \( (Ex)PaFx \) holds, then two other formulas hold, \( PaFb \) and \( (Ex)Ka(b=x) \) for some constant \( b \). And this is entirely unreasonable. For it says that if there is someone who \( a \) does not know to be non-\( F \), then \( a \) has a unique way of referring to that person. Since constant terms have descriptive content for Hintikka, this means that if there is someone of whom you are ignorant in some respect, then you know something unique about him. But I believe, by our ordinary understanding there can be individuals of whom one knows nothing at all. 5

And according to Sleigh:

What \( E(108) \) seems to say when applied to \( (Ex)-KaFx \) is that anyone \( a \) doesn't know to be \( F \) is among the persons known to \( a \). This seems obviously unacceptable. 6

There is no doubt that the inference

\[
\begin{align*}
(Ex)PaFx & \in w \\
(Ex)Ka(b=x) & \in w, \text{ for some } 'b'
\end{align*}
\]

(let's call it "(*)") is highly counterintuitive. But notice that the proof of (a) does not rely upon this particular inference. A close examination of the above proof shows that step (d) is a wet noodle; erase it and the proof is not affected. Likewise, proofs of (c) and (d) do not require (*); it is only when we try to prove (b) that this inference is needed to reach a contradiction.

The other consequence of \( (Ex)PaFx \) licensed by (108), \( 'PaFb' \), is not (as far as English readings go) counterintuitive. But it is this inference (let's call it "(**)") which is of concern in the present context of discussion, for all of the formulas (a)-(d)
depend upon it to be proven in KB. It is the crucial move by which one is enabled to get rid of the quantifier(s) in order to eventually reach the contradiction buried inside the scope of 'Pa'.

Therefore, without (**) none of the formulas in question can be proven in KB. And it is precisely this same inference which makes (a)-(d) theorems in Tienson's F* as well. There one is allowed to infer \( \lambda(x/x)' \in M \) from \((Ex)PaAx' \in M\), for some auxiliary symbol \( \lambda \). And in Sleigh's K, although the whole of proof procedure is couched in terms of interpretation functions, the result is the same. If \( V_1((Ex)PaFx, w_1) = l \), then for some \( I' \) which is like \( I \) except (possibly) at \( x \), \( I'(x,w_1) \in Di \) and \( V_{I'}(PaFx,w_1) = l \). There is an assignment of \( x \) to an element in \( w_1 \)'s domain which makes 'PaFx' true in \( w_1 \). Thus one can peel off the quantifier just as in KB and F* and eventually reach the contradictions which suffice to establish (a)-(d) as theorems.

In order, therefore, to find a way to block proofs of (a)-(d), it looks as though rules for instantiation of \((Ex)PaFx\) must be pretty radically revised. According to Sleigh:

What is required is one set of instantiating conditions for formulas of the form \((Ex)Ka\emptyset\) and a distinct set for formulas of the form \((Ex)Pa\emptyset\). It is possible to develop a tenable model set approach along these lines which is free of the restricted range feature of KB but the resulting system is surprisingly complicated.

A way must be found to grant instantiation upon \((Ex)PaFx\) only under certain specified conditions and not in the same indiscriminate fashion accorded to \((Ex)KaFx\). Can such a technique be found? And would such a technique, if it could be found, make good sense not only on a formal level but relative to an intuitive possible worlds interpretation as well?
I shall now proceed to propose a system of model sets \( KB^* \) involving just such a technique.

Consider the following quantifier rules for contexts in which variables occurring inside the scope of a single epistemic operator are bound from the outside by quantifiers:

\[
\begin{align*}
(\text{EX}) & \quad (\exists x)K_aF_x \leq w \\
& \quad (\exists x)K_a(b=x) \leq w \\
& \quad KaF_b \leq w, \text{ for some } 'b' \\
(\text{UK}) & \quad (x)K_aF_x \leq w \\
& \quad (\exists x)(b=x) \leq w \\
& \quad (\exists x)(b=x \land KaF_x) \leq w \\
(\text{EPL}) & \quad (\exists x)P_aF_x \leq w \\
& \quad (\exists x)(b=x) \leq w \\
& \quad (\exists x)(b=x \land P_aF_x) \leq w, \text{ for some } 'b' \\
(\text{EPS}) & \quad (\exists x)P_aF_x \leq w \\
& \quad (x)K_aG_x \leq w \\
& \quad P_aF_b \leq w, \text{ for some } 'b' \\
(\text{EPM}) & \quad (\exists x)(G_x \land P_aF_x) \leq w \\
& \quad (x)K_aH_x \leq w \\
& \quad G_c \land P_aF_b \leq w, \text{ for some } 'b', 'c' \\
(\text{UPL}) & \quad (x)P_aF_x \leq w \\
& \quad (\exists x)(b=x) \leq w \\
& \quad (\exists x)(b=x \land P_aF_x) \leq w \\
(\text{UPS}) & \quad (x)P_aF_x \leq w \\
& \quad (\exists x)K_a(b=x) \leq w \\
& \quad P_aF_b \leq w.
\end{align*}
\]

We shall now form \( KB^* \) in the following way. Begin with the rules of \( KB \). Replace (103) and (109) by the seven rules listed above. Drop from \( KB \) the rules

\[
\begin{align*}
(\text{C.EK.}^*) & \quad (\exists x)K_a(b=x) \leq w \\
& \quad Ka(\exists x)(b=x) \leq w \\
(\text{C.EK.EK.}^*) & \quad (\exists x)K_a(b=x) \leq w \\
& \quad (\exists x)K_a(b=x) \leq w^*, \text{ for every } w^* \text{ such that } Rw^*.
\end{align*}
\]

Add the rules
And to all of this we add a set of rules, to be described in what follows, to govern formulas formed by quantifying in contexts two or more layers deep of epistemic operators.

Our first order of business is to establish that our system lacks the Restricted Range feature. For this it suffices to show that

\[
\frac{(x)KaFx}{(x)Fx} \in W
\]

is provable in KB*:

\[
\begin{align*}
(a) \quad & (x)KaFx \quad \varepsilon W \quad \text{Redog} \\
(b) \quad & -(x)Fx \quad \varepsilon W \\
(c) \quad & (Ex)Fx \quad \varepsilon W \quad (b), (c.-U) \\
(d) \quad & -Fb \quad \varepsilon W \quad (c), (C.E_0), \text{ for some } 'b' \\
(e) \quad & (Ex)(b=x) \quad \varepsilon W \quad (c), (C.E_0), \text{ for some } 'b' \\
(f) \quad & (Ex)(b=x \land KaFx) \varepsilon W \quad (a), (e), (UK) \\
(g) \quad & b=c \land KaFc \quad \varepsilon W \quad (f), (EK) \\
(h) \quad & (Ex)Ka(c=x) \quad \varepsilon W \quad (g), \text{Simp} \\
(i) \quad & KaFc \quad \varepsilon W \quad (i), (C.K*), \text{Reflexivity} \\
*(j) \quad & Fc \quad \varepsilon W
\end{align*}
\]
Notice that a similar proof cannot be produced in KB; one may proceed as far as step (e), but the move from (a) and (e) to (f) is not allowed in KB.

As far as instantiation on \((x)KaFx\) is concerned, our system is actually more liberal than KB. Our rule (UK) allows instantiation on \((x)KaFx\), in a sort of large-scope fashion, with the auxiliary clause \((Ex)(b=x)\); KB has no such provision. At the same time, Hintikka's (small-scope) instantiation rule

\[
\frac{(x)KaFx}{(Ex)Ka(b=x)} \in W
\]

is a consequence of our rules:

- (a) \((x)KaFx \in W\)
- (b) \((Ex)Ka(b=x) \in W\)\{Redog
- (c) \(-KaFb \in W\)
- (d) \((Ex)(b=x) \in W\)\{Redog, (C.Ind.=n)
- (e) \((Ex)(b=x & KaFx) \in W\)\{Redog, (UK)
- (f) \(b=c & KaFc \in W\)\{Redog, (EK)
- (g) \((Ex)Ka(c=x) \in W\)\{Redog, (EK)
- *(h) \(KaFc \in W\)\{Redog, (E), (EK)
- (i) \(b=c \in W\)\{Redog, (E), (EK)
- (j) \(Ka(b=c) \in W\)\{Redog, (E), (EK)
- *(k) \(-KaFc \in W\)\{Redog, (E), (EK)

At the other end of the spectrum, however, our instantiation rules for 'Pa' are much less liberal than in KB. This, of course, is all a part of our attempt to block proofs of formulas (a)-(d)
and (a')-(d'). If we try, for example, to prove (a) by assuming that \( (\text{Ex})\overline{\text{Pa}(x\neq x)} \) belongs to a model set \( w \), we can infer both \( (\text{Ex})(b=x)' \in w \) and \( (\text{Ex})(b=x \& \text{Pa}(x\neq x))' \in w \). But we can infer nothing that yields a contradiction; we lack the means to peel off the quantifier and get at what's inside the epistemic operator.

Thus our rules succeed very nicely in blocking the proofs of the unwanted theorems by making at least one of Hintikka's key quantifier rules drastically weaker. But an important objection arises here. Might not the weakening of (108) be an ad hoc formal maneuver which fails to make good sense relative to an intended possible worlds interpretation of formulas in KB*? Perhaps the natural intuitions which underlie Hintikka's rules taken as a whole will be, once these rules have been tampered with, gone for good and we shall be left with a new set of rules which are artificial, contrived, and unable to be backed with the authority of an intuitive model.

In reply to this objection we now proceed to the arduous task of showing that while it is true that the intuitions underlying Hintikka's rules are gone for good, we can show that underlying our rules is a set of intuitions every bit as natural as Hintikka's. Therefore, let us turn our attention towards a possible worlds interpretation \( M \) of the formulas in KB* and examine our rules in the light of this interpretation. Rather than simply describe our model \( M \), however, we shall follow Sleigh's lead and construct an interpreted, higher-order theory \( W \) which "talks about" \( M \) in that the truth-conditions for formulas of KB* relative to \( M \) are made explicit. In this way we can judge with a good deal more precision the adequacy of the rules of KB*. Following Sleigh, we shall call \( W \) the "world
theory" of KB*, and a formula in W which describes the truth-conditions in M of a formula 'A'. Having done this, it will be very easy to judge whether KB* is an adequate system. Any inference

\[
\begin{array}{c}
\text{A} \\
\text{B}
\end{array} \in W
\]

must be provable in KB* if and only if \( 'A* \rightarrow B* ' \) is provable in W, where \( 'A*' \) and \( 'B*' \) are the respective world theory transcriptions of 'A' and 'B'.

The theory we shall utilize is in its essentials due to E. Gettier. The language of W can be described as follows:

(i) Every symbol in KB* except 'Ka' and 'Pa' belongs to the language of W;

(ii) In addition:

'\text{I*}' is a one-place predicate constant in W

'E' is a two-place predicate constant in W

'R' is a two-place predicate constant in W

'B_1', 'B_2', etc. are two-place predicate constants in W

'P_1', 'P_2', etc. are two-place predicate variables in W

'w_{i}\', w_{j}\', etc. are variables in W

'w_r' is a constant in W;

(iii) Atomic well-formed formulas are as follows:

(a) If 'A' is a wff in KB*, then if 'A' contains no occurrences of 'Ka', 'Pa', or individual constant terms, \( '(A)w_i' \) is a wff in W, for any i

(b) \( 'I^*P_i' \) is a wff in W, for any i

(c) \( 'Ewx_i' \) is a wff in W, for any i and any variable or constant x in W

(d) \( 'Rw_iw_j' \) is a wff in W, for any i, j
(e) 'P_i x w_j' is a wff in W, for any i, j and any variable or constant x in W
(f) 'B_i x w_j' is a wff in W, for any i, j and any variable or constant x in W

Intuitively, atomic formulas in W are to be understood in the following way. '(A)_w_i' asserts that a formula 'A' of KB* is true in world w_i (relative to M). 'I*_P_1' asserts that a world-line P_1 is 'privileged' in the sense that quantifiers in KB* reaching into epistemic contexts range over a set of world-lines of which P_1 is a member. 'Ex_w_i' is understood to say that x exists in world w_i. 'P_i x w_j' asserts that world-line P_1 picks up object x at world w_j. 'R_w_i x w_j' asserts that world w_j is compatible with world w_i. And 'B_i x w_j' is understood to say simply that x is B_i at w_j. In what follows we shall refer to the 'B_i' predicates as "Badge Predicates".

I believe we now have the machinery to express in the language of W the explicit truth-conditions of any formula in KB* relative to our intended model M. We now give a recursive procedure to show how to derive, for any formula 'A' in KB*, the world theory transcription 'A*' of 'A' in W. We begin by defining the notion 'Immediate W-translation' ('IW-translation' for short). For purposes of abbreviation we shall shorten 'w_j' to 'j' throughout, and for convenience we shall assume that 'I' and '-' are the only truth-connectives in 'A'.

(i) If F is atomic, the IW-translation of F is '(F)i'
(ii) If F is of the form 'B & C', the IW-translation of F is '(B)i & (C)i'
(iii) If F is of the form '-B', the IW-translation of F is '-((B)i)'
If \( F \) is of the form \( \text{'KaB} \) then

(a) If every free variable in \( B \) is bound in \( B \) by at least one epistemic operator, the IW-translation of \( F \) is \( '(j)(Rij \rightarrow (E_ij) 'B)j' \)

(b) If \( x_1, \ldots, x_n \) occur free in \( B \) outside the scope of all epistemic operators in \( B \), the IW-translation of \( F \) is \( '(j)(Rij \rightarrow (E\{y\} \rightarrow (P^v_{k-1} \cdot \ldots \cdot P^v_1 y_{k-1} \cdot \ldots \cdot P^v_n y_n \cdot (B(y/x)))j) ' \), where \( F \) occurs within the context \( ' \ldots (EP_1)((I^*P_1 \cdot P_1 x_1) \cdot \ldots (EP_n)((I^*P_n \cdot P_n x_n) \cdot \ldots F \ldots) \ldots) ' \)

(v) If \( F \) is of the form \( \text{'PaB} \), the IW-translation of \( F \) is \( '-\text{Ka-B} ' \)

(vi) If \( F \) is of the form \( '\text{'(Ex)B} ' \), then

(a) If \( 'x' \) doesn't occur free in \( B \), the IW-translation of \( F \) is \( '(B)i' \)

(b) If \( 'x' \) occurs free in \( B \) but never within the scope of an epistemic operator, the IW-translation of \( F \) is \( '(\text{Ex})(\text{Exi} \cdot (B)i)' \)

(c) If \( 'x' \) occurs free in \( B \) and for at least one occurrence of \( 'Ka' \) not directly preceded by an odd number of negation signs, \( 'x' \) occurs within its scope and it (the \( 'Ka' \)) occurs within the scope of no epistemic operator, then

(i) If \( B \) is of the form \( 'C \& D' \) or \( '-(C \& D)' \) and \( 'x' \) occurs only within the scope of epistemic operators, then the IW-translation of \( F \) is \( '(\text{Ex})(\text{Exi} \cdot (EP_1) (I^*P_1 \cdot P_1 x_i \& (C)i) \& (EP_2)(I^*P_2 \cdot P_2 x_i \& (D)i))' \) or \( '(\text{Ex})(\text{Exi} \& -((EP_1)(I^*P_1 \cdot P_1 x_i \& (C)i) \& (EP_2)(I^*P_2 \cdot P_2 x_i \& (D)i)))' \), respectively
(ii) Otherwise the IW-translation of $F$ is 
$$(\text{Ex})(\text{Ex}i \& (\text{EP}_1)(I^*P_1 \& P_1x_i \& (B)i))$$.

(d) In all other cases the IW-translation of $F$ is 
$$(\text{Ex})(\text{Ex}i \& (P^^Kd^P^ \& P^jx_i)-(\text{B})i)$$.

(viii) If $F$ is of the form 
$$'(x)B'$$,
the IW-translation of $F$ is 
$$'-(\text{Ex})-B'$$.

Here we let '$i$' be a meta-linguistic variable whose value is determined as follows. If '$F$' is '$A'$, the '$i$' stands for '$r$'. And if '$F$' is '$E'$ in the procedure described below, then '$i$' stands for the variable '$j$' in the quantifier '$(E)j$' or '$(j)$' most directly binding '$E$'. If '$F$' is '$E'$ and there is no such quantifier, then '$i$' stands for '$r$'.

We now say that a formula '$G$' is the 'W-translation' of a formula '$F$' if and only if

(i) There is a finite sequence of formulas $F_1, \ldots, F_n$ such that $F_1$ is the IW-translation of $F$, $F_{i+1}$ is the IW-translation of $F_i$, for all $i$, and $G$ is the IW-translation of $F_n$, and

(ii) For every subformula in $G$ of the form '$(A)i$', '$A$' is atomic. On the basis of all of this we now calculate '$A^*$' in the following way.

Let '$A_1$' be the Immediate W-translation of '$A$'. Now let '$D$' be the leftmost subformula in '$A_j$' which is of the form '$(E)i$' for some $i$ and non-atomic formula '$E$'. Then '$A_{j+1}$' is the formula obtained by replacing '$D$' by the W-translation of '$E$', if it has one, and by its Immediate W-translation otherwise. When a formula '$A_j$' is reached in which there is no such subformula '$D$', the '$A_j$' is '$A^*$'.

Let us now look at some examples of basic formulas and their respective world theory transcriptions in $W$. From now on we shall often abbreviate '$(B)i$' as '$Bi$'.

(i) KaB: 
$$(i)(Fri \rightarrow (B)i)$$.
(ii) \((\text{Ex})\text{KaBx}:\)
\((\text{Ex})(\text{Exr} & (\text{EP}_1)((I^*\text{P}_1 & \text{P}_1x\text{r}) & (i)(\text{Rri}\rightarrow(\text{Ey})(\text{P}_1\text{yi} & \text{Eyi}))))\)

(iii) \((x)\text{KaBx}:\)
\((x)(\text{Exr} \rightarrow(\text{EP}_1)((I^*\text{P}_1 & \text{P}_1x\text{r}) & (i)(\text{Rri}\rightarrow(\text{Ey})(\text{P}_1\text{yi} & \text{Eyi}))))\)

(iv) \((\text{Ex})\text{Cx} \& \text{KaBx}:\)
\((\text{Ex})(\text{Exr} & (\text{EP}_1)((I^*\text{P}_1 & \text{P}_1x\text{r}) & \text{Cxr} & (i)(\text{Rri}\rightarrow(\text{Ey})(\text{P}_1\text{yi} & \text{Eyi}))))\)

(v) \((\text{Ex})\text{PaBx}:\)
\((\text{Ex})(\text{Exr} & (\text{P}_1)((I^*\text{P}_1 & \text{P}_1x\text{r}) \rightarrow(\text{Ei})(\text{Rri} & (\text{Ey})(\text{P}_1\text{yi} & \text{Eyi}))))\)

(vi) \((\text{Ex})\text{KaBx} \& \text{PaCx})\)
\((\text{Ex})(\text{Exr} & (\text{EP}_1)((I^*\text{P}_1 & \text{P}_1x\text{r}) & ((i)(\text{Rri}\rightarrow(\text{Ey})(\text{P}_1\text{yi} & \text{Eyi})) & (\text{Ei})(\text{Rri} & (\text{Ey})(\text{P}_1\text{yi} & \text{Cyi}))))\)

(vii) \((\text{Ex})\text{KaKaBx}:\)
\((\text{Ex})(\text{Exr} & (\text{EP}_1)((I^*\text{P}_1 & \text{P}_1x\text{r}) & (i)(\text{Rri}\rightarrow(j)(\text{Rij}\rightarrow(\text{Ey})(\text{P}_1\text{yl} & \text{Byj}))))\))

(viii) \((\text{Ex})\text{KakaBx})\)
\((\text{Ex})(\text{Exr} & (\text{EP}_1)((I^*\text{P}_1 & \text{P}_1x\text{r}) & (i)(\text{Rri}\rightarrow((\text{Ey})(\text{P}_1\text{yi} & \text{Eyi})) & (j)(\text{Rij}\rightarrow(\text{Ey})(\text{P}_1\text{yl} & \text{Byj}))))\))

(ix) \((\text{Ex})\text{PaKaBx}:\)
\((\text{Ex})(\text{Exr} & (\text{P}_1)((I^*\text{P}_1 & \text{P}_1x\text{r}) \rightarrow(\text{Ei})(\text{Rri} & (j)(\text{Rij}\rightarrow(\text{Ey})(\text{P}_1\text{yl} & \text{Byj}))))\))

Before moving on, there is one technical difficulty which remains to be corrected in the procedure we have just presented. Constant singular terms in KB* are capable of referring to different individuals in different worlds, but constant singular terms in 𝑊 have a fixed reference (always designate rigidly). Therefore, to make sure wobbly terms in KB* do not metamorphize into rigid terms in the course of our procedure, we must make the following revision.

If 'B' is a formula in KB* having no constant singular terms (as in all of the above examples) let 'B' be 'A' in the above procedure. If 'B' contains constant singular terms we derive 'A' as follows. Since KB* is a system of our own making there is no problem in assuming that for each constant singular term 'b' in KB*, there is
a badge predicate in KB* which holds uniquely of b. Replace every subformula '((Ex)C)' in 'B' by '(Ex)(Ex & C)', every subformula '(x)C' in 'B' by '(x)(Ex→C)', and every atomic subformula 'D' in 'B' which contains a constant singular term 'b' by '(Ex)(D(b/x) & B_yx)', where 'B_yj' is the badge predicate of 'b' in KB* and 'E' is not a symbol in KB*. When all singular terms have been eliminated the result is 'A'. Now go to the recursive procedure outlined above with the following revision of 'IW-translation'. Add to the first line of (vi), "...if B is of the form 'Ex & C', then", and add "(vii) If F is of the form '(Ex)B' and B is not of the form 'Ex & C', then..." followed by (a)-(d) of (vi) with the clause 'Exi' deleted everywhere it occurs.

It is a simple matter, incidentally, to assign world-theory transcriptions in W to formulas in Hintikka's KB. We need to make only two alterations to our definition of 'IW-translation' as follows: Replace '((j)(R_yj→(Ey_1)...(Ey_n)P_y_j & ... & P_y_j & B(y_k/x_k)))' by '((j)(R_yj→(Ey_1)...(Ey_n)P_y_j & ... & P_y_j & (Ey_j & ... & Ey_j) & (B(y_k/x_k))))', and delete part (vi)-(c).

The differences in the intuitions underlying the rules of KB and those underlying the rules of KB*, therefore, are twofold. Privileged world-lines in KB pick out individuals in worlds only if they happen to exist in such worlds. And the truth of '(x)KaFx' in a world does not in KB necessitate that every existent individual in that world is picked up by a privileged world-line, as is the case in KB*. It is hard to see that either set of intuitions is in any profound way simpler or more natural than the other.

So far we have spoken only of the language of W. In order to
tell whether our rules for $KB^*$ are adequate, however, we must specify the axioms and rules of inference of $W$. As far as the logical axioms and rules of $W$ are concerned, we can simply treat $W$ as a second-order theory ($W$ having no modal operators). What requires specification is the set of non-logical axioms of $W$. These shall be as follows:

(i) $(i)(Rii)$

(ii) $(i)(j)(k)((Rij \& Rjk) \rightarrow Rik)$

(iii) $(P_1)(i)(Ex)(P_1xi \& (y)(P_1yi \equiv P_1xi))$

(iv) $(P_1)(P_2)(x)(y)(i)(j)((P_1xi \& P_2xi \& Rij) \rightarrow (P_1yj \equiv P_2yj))$

(v) $[(EP_1)(i)(x)(P_1xi \equiv (Qx)i)]$ is an axiom, for any badge predicate '$\alpha$'.

Let us now take a closer look at our system $KB^*$ through the eyes of $W$. We look first to see how the issue of Restricted Range is handled relative to world theory transcriptions in $W$. In English the world theory transcription of '$(x)KaFx$' (see above) may be read, " Anything which exists in the actual world is picked up by a privileged world-line which in every compatible world picks up something which is $F$ in that world". The world theory transcription in $W$ of the same formula in $KB$, on the other hand, reads, " Anything which exists in the actual world and which is picked up by a privileged world-line is picked up by a world-line which in every compatible world picks up something which exists and is $F$ in that world". Thus while '$(x)KaFx$' $\in w$ holds in $KB^*$ only if every individual in $w$'s domain is picked up by a privileged world-line, only a restricted class of $w$'s domain needs to be picked up by a privileged world-line for the same condition to hold in $KB$. In this way the
lack of a Restricted Range feature in KB* is made explicit through
the world theory transcriptions assigned to formulas in KB*.

To verify that this is indeed the case we now show that the
demonstration produced above to show that '(x)KaFx' virtually implies
'(x)Fx' can be paralleled in W.13

\[ (a) \quad (x)(Exr \rightarrow (EP_1)((I*P_1 \& P_1xr) & (i)(Rri \rightarrow (Ey)(P_1y_i \& Fyi)))) \]
\[ (b) \quad Ebr \rightarrow (EP_1)((I*P_1 \& P_1br) & (i)(Rri \rightarrow (Ey)(P_1y_i \& Fyi))) \]
\[ (c) \quad Ebr \rightarrow (EP_1)((I*P_1 \& P_1br) & (i)(Rri \rightarrow (Ey)(P_1y_i \& Fyi))) \]
\[ (d) \quad (EP_1)((I*P_1 \& P_1br) & (i)(Rri \rightarrow (Ey)(P_1y_i \& Fyi))) \]
\[ (e) \quad (I*P & Pbr) & (i)(Rri \rightarrow (Ey)(Fyi & Fyi)) \]
\[ (f) \quad (i)(Rri \rightarrow (Ey)(Fyi & Fyi)) \]
\[ (g) \quad Rrr \rightarrow (Ey)(Pyr & Fyr) \]
\[ (h) \quad Rrr \rightarrow (Ey)(Pyr & Fyr) \]
\[ (i) \quad (Ey)(Pyr & Fyr) \]
\[ (j) \quad Pcr & Fcr \]
\[ (k) \quad I*P & Pbr \]
\[ (l) \quad Pbr \]
\[ (m) \quad Pcr \]
\[ (n) \quad (b=c)r \]
\[ (o) \quad Fcr \]
\[ (p) \quad Fbr \]
\[ (q) \quad Fbr \]
\[ (r) \quad Fbr \]
\[ (s) \quad Ebr \rightarrow Fbr \]
\[ (t) \quad (x)(Exr \rightarrow Fxr) \]

We are now in a position to justify our instantiation rules for
'(Ex)PaFx'. We first argue that (**) ought not be provable in KB*.
We assume \( (Ex)(Exr \& (P_1)((I*P_1 \& P_{1xr})\rightarrow (Ei)(Rri \& (Ey)(P_{1yi} \& Fyi)))) \) and try to deduce \( (Ei)(Rri \& (Ey)(Fyi \& Bjyi)) \), for some 'Bj'.

\[
\begin{align*}
(a) & \quad (Ex)(Exr \& (P_1)((I*P_1 \& P_{1xr})\rightarrow (Ei)(Rri \& (Ey)(P_{1yi} \& Fyi)))) \\
(b) & \quad (Ebr \& (P_1)((I*P_1 \& P_{1br})\rightarrow (Ei)(Rri \& (Ey)(P_{1yi} \& Fyi)))) \\
(c) & \quad (P_1)((I*P_1 \& P_{1br})\rightarrow (Ei)(Rri \& (Ey)(P_{1yi} \& Fyi))) \quad \text{Ass.} \\
(d) & \quad (I*P \& P_{br})\rightarrow (Ei)(Rri \& (Ey)(P_{yr} \& Fyi)) \quad \text{c), UI}
\end{align*}
\]

Here we run stuck. To obtain \( (Ei)(Rri \& (Ey)(Fyi \& Bjyi)) \) it is absolutely essential to fish out the 'Rri' and 'Fyr' clauses in the consequent of step (d), but there is no way on earth to get at these. And so the proof is blocked.

In like fashion the inference (*) cannot be backed up by the world theory. There is no way to infer \( (EP_1)((I*P_1 \& P_{1br}) \& (i) (Rri\rightarrow (Ey)(P_{1yi} \& (Ea((y=a)i \& Bjai)))) \) from what we have in step (d) above. Neither (d) nor any step which precedes it guarantees that some privileged world-line actually does pick up b at the real world. Hence there is no hope in reaching the world theory transcription of \( (Ex)Ka(b=x)' \); the proof is blocked.\(^\text{14}\)

On the other hand, the two instantiation rules for \( (Ex)PaFx' \) we introduce in KB*, (EPL) and (EPS), are justified in W. The case of (EPL) is obvious; it is only a question of showing that from \( (Ex)(Exr \& (P_1)((I*P_1 \& P_{1xr})\rightarrow (Ei)(Rri \& (Ey)(P_{1yi} \& Fyi)))) \) we can deduce \( (Ex)(Exr \& (P_1)((I*P_1 \& P_{1xr})\rightarrow ((Ey)((y=x)r \& Bjxr) \& (Ei) (Rri \& (Ey)(P_{1yi} \& Fyi)))))) \), for some 'Bj', and \( (Ex)(Exr \& (Ey)((y=x)r \& Bjyr)) \), for some 'Bj'. And both of these are easily proven.

The proof of (EPS) goes as follows:
(a) \((\exists x)(\exists y)((P_1 \land P_1 \rightarrow (Ei)(Rr \land (Ey)(P_1 y \land F y)))\)

(b) \((\exists x)(\exists y)((P_1 \rightarrow ((Ei)(Rr \land (Ey)(P_1 y \land G y)))\)

(c) \((\exists c)(\exists P_1)((P_1 \land P_1 \rightarrow (Ei)(Rr \land (Ey)(P_1 y \land F y)))\)

(d) \((\exists c)(\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow ((Ei)(P_1 y \land G y)))\)

(e) \((\exists c)(\exists P_1)\)

(f) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(g) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(h) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(i) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(j) \((\exists P_1)\)

(k) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(l) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(m) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(n) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(o) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(p) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(q) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(r) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(s) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(t) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(u) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(v) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(w) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(x) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)

(y) \((\exists P_1)((P_1 \land P_1 \rightarrow (i)(Rr \rightarrow (Ey)(P_1 y \land G y)))\)
It is now apparent that blocking the proofs of formulas (a)-(d) by rejecting (*) and (**) in favor of alternative rules which are substantially weaker is not an ad hoc syntactic device. Rather, it is a move backed up by the full authority of an intuitive possible worlds interpretation of the formulas in KB* as made explicit in the world theory W.

According to this particular interpretation, '(x)KaFx' can be true in a world w only if every individual in the domain of w is attached to a world-line over which '(x)' ranges. And so we understand '(x)KaFx' as asserting, for any x in the domain of w, not only that every privileged world-line picks up something in every compatible world which in that world is F, but that x is picked up by exactly one of these world-lines. Therefore, '(x)-Ka-Fx' must be regarded as true in w just in case for any x in the domain of w either not every privileged world-line in every compatible world picks up something which is not-F or x is not picked up by one of these world-lines. And this gives us precisely the world theory transcription of '(x)PaFx' generated by our recursive procedure. By the same token, '(Ex)PaFx' is true in w just in case for some x in the domain of w not every privileged world-line in every compatible world picks up something which is not-F or else x is not picked up by one of these world-lines. Given this interpretation it is easy to see why (*) and (**) fail in KB* and hence why it is fitting that (a)-(d) should not be provable in KB*.

It is at this point that Stine's English readings for epistemic formulas become instructive. Given that an individual x in the domain of w can be said to be "known to a" in w just in case one of
a' s privileged world-lines picks up x in w, it is now apparent why
Stine urges that '(x)KaFx' be read

(5) Everyone is known to a and known by a to be F.

It is understood to make a two-fold (conjunctive) assertion about
each individual x in the domain of the world in question. Consequently,
'(x)PaFx' must be understood to make a disjunctive assertion about
each such individual, and hence it is read

(6) Everyone is either not known to a or not known by a to be not-F,
which is equivalent to

(7) No one is known to a and known by a to be not-F.

In the same way, '(Ex)KaFx' makes a conjunctive assertion about
a single individual, and this fact is reflected in Stine's reading for
'(Ex)KaFx', viz.

(8) Someone is known to a and known by a to be F.

And by the same token, '(Ex)PaFx' makes a disjunctive assertion about
a single individual and is read

(9) Someone is either not known to a or not known by a to be not-F.

Therefore, Stine's readings for '(x)KaFx', '(x)PaFx', '(Ex)KaFx',
and '(Ex)PaFx', which are (respectively) (5), (7), (8), and (9),
parallel the world theory transcriptions of these formulas which we
have introduced.

It is true, nonetheless, that while Stine lists (5) as a reading
for '(x)KaFx' and (8) as a reading for '(Ex)KaFx', '(x)PaFx' is
actually listed as having as a reading not (7) but rather

(10) No one is known by a to be not-F,

and '(Ex)PaFx' is officially given as a reading

(11) Someone is not known by a to be not-F
rather than (9). As we pointed out in the last chapter, however, if we accept the truth of

(12) Everyone known by a to be F is known to a
(as Stine does), then it is easily proven that (7) is equivalent to (10) and (9) is equivalent to (11). Moreover, if we assume (12), it can be shown that (5) is equivalent to

(13) Everyone is known by a to be F
and (8) is equivalent to

(14) Someone is known by a to be F.

Therefore, not only are we justified in reading '(x)KaFx' as (5), '(x)PaFx' as (7), '(Ex)KaFx' as (8), and '(Ex)PaFx' as (9). But if we accept (12) we may actually go a step further and adopt the following much simpler English lexicon:

(x)KaFx - "Everybody is known by a to be F"
(Ex)KaFx - "Someone is known by a to be F"
(x)PaFx - "No One is known by a to be F"
(Ex)PaFx - "Someone is not known by a to be not-F".

It would be very nice to adopt this set of English readings. But there is no justification in doing so until we can first justify accepting the truth of (12). To do this it suffices to show that the world theory transcription in W of the formal rendition of (12) in KB* (which is Stine's axiom, '(x)(KaFx→Ka(x=x))') is a theorem in W. And this can be done as follows:

\[
\begin{align*}
&\vdash (a) \text{ Ebr} \\
&\vdash (b) (EP_1)((I*P_1 \land P_1 \text{ br}) \land (i)(\text{Rri→}(Ey)(P_1 y_i \land F y_i))) \text{ Ass.} \\
&\vdash (c) (I*P \land P \text{ br}) \land (i)(\text{Rri→}(Ey)(P y_i \land F y_i)) \text{ Ass.} \\
&\vdash (d) (i)(\text{Rri→}(Ey)(P y_i \land F y_i)) \text{ (c), Simp}
\end{align*}
\]
So far we have introduced a world theory \( W \) on the basis of which the set of rules of our system \( KB^* \) may be tested for adequacy. We have discovered that at the crucial point at which our rules diverge...
from those of KB (as well as F* and K), the instantiation rules for
'(Ex)PaFx', W lends full support to the rules of KB*. Thus we have
not only a system in which proofs of (a)-(d) are blocked, but we
have a system whose means of blocking these proofs has the full
authority of a natural, intuitive possible worlds interpretation of
the formulas of KB*. Finally, we have seen that this same inter-
pretation allows us to adopt a set of English readings for formulas
in KB* just like those introduced by Stine for her system.

Yet so far we have said nothing about rules for KB* governing
contexts in which variables within the scope of two or more epistemic
operators are bound from the outside by quantifiers. We give first
a complete set of rules covering cases which involve two epistemic
operators. Then on the basis of these together with the original
seven quantifier rules we listed, we shall propose a concise set of
rules general enough to cover all cases in which variables occur
within epistemic operators and are then bound from the outside. The
resulting set of rules we shall officially adopt as our quantifier
rules.

The following is a complete set of rules governing formulas
formed by quantifying in past two epistemic operators:

\[(EKK) \begin{align*}
(Ex)KaKaFx & \in w \\
(Ex)KaKa(b=x) & \in w \\
KaKaFb & \in w \\
(Ex)KaFx & \in w, \text{ for some } 'b' \\
\end{align*}\]

\[(UKK) \begin{align*}
(x)KaKaFx & \in w \\
(Ex)(b=x) & \in w \\
(Ex)(b=x & KaKaFx) & \in w \\
(x)KaFx & \in w \\
\end{align*}\]

\[(EPPL) \begin{align*}
(Ex)PaPaFx & \in w \\
(Ex)(b=x) & \in w \\
(Ex)(b=x & PaPaFx) & \in w \\
(Ex)PaFx & \in w, \text{ for some } 'b' \\
\end{align*}\]
In applying rules (EPPS), (EKP), and (EPKS) the following qualification must be observed. In any clause bound by '(Ex)' which contains free 'x' but fails to fall within the scope of two epistemic operators the 'x' may be instantiated upon but not by the constant already chosen for free 'x's' which do fall into the scope of two epistemic operators. A new constant must be used for each such clause; two clauses, however, with the same modal profile may share constants. By way of example:
Now let 'Qa' and 'Qi'a' be short for arbitrary (Possibly empty) strings formed by concatenating 'Ka' and 'Pa'. We may then formulate the official version of our quantifier rules for KB* as follows:

\[
\begin{align*}
& (\text{EX}) & (\text{Ex})KaQaFx & \leq w \\
& & (\text{Ex})KaQa(b=x) & \leq w \\
& & KaQaFb & \leq w, \text{ for some } 'b' \\
& (\text{EKM}) & (\text{Ex})KaQa(Fx & Q'i'aGx) & \leq w \\
& & KaQa(Fc & Q'i'aGb) & \leq w \\
& & (\text{Ex})KaQa(c=x & Q'i'a(b=x)) & \leq w, \text{ for some } 'c', 'b' \\
& (\text{UK}) & (x)KaQaFx & \leq w \\
& & (\text{Ex})(b=x) & \leq w \\
& & (\text{Ex})(b=x & KaQaFx) & \leq w \\
& (\text{EPL}) & (\text{Ex})PaQaFx & \leq w \\
& & (\text{Ex})(b=x) & \leq w \\
& & (\text{Ex})(b=x & PaQaFx) & \leq w, \text{ for some } 'b' \\
& (\text{EPS}) & (\text{Ex})PaQaFx & \leq w \\
& & (x)KaGx & \leq w \\
& & PaQaFb & \leq w, \text{ for some } 'b' \\
& (\text{EPM}) & (\text{Ex})PaQa(Fx & Q'i'aGx) & \leq w \\
& & (x)KaGx & \leq w \\
& & PaQa(Fc & Q'i'aGb) & \leq w, \text{ for some } 'c', 'b' \\
& (\text{UPL}) & (x)PaQaFx & \leq w \\
& & (\text{Ex})(b=x) & \leq w \\
& & (\text{Ex})(b=x & PaQaFx) & \leq w \\
& (\text{UPL}) & (x)PaQaFx & \leq w \\
& & (\text{Ex})Ka(b=x) & \leq w \\
& & PaQaFb & \leq w \\
& (\text{EPE}) & (\text{Ex})QaPaPaQ'i'aFx & \leq w \\
& & (\text{Ex})QaPaQ'i'aFx & \leq w \\
& (\text{EKE}) & (\text{Ex})QaKaQ'i'aFx & \leq w \\
& & (\text{Ex})QaQ'i'aFx & \leq w \\
& (\text{EKI}) & (\text{Ex})QaKaQ'i'aFx & \leq w \\
& & (\text{Ex})QaKaQ'i'aFx & \leq w \\
& (\text{EPI}) & (\text{Ex})QaPaQ'i'aFx & \leq w \\
& & (\text{Ex})QaPaPaQ'i'aFx & \leq w \\
\end{align*}
\]
Rules (EKM) and (EPM) are to be treated as qualifications to (EK) and (EPS), respectively. The latter may be applied only under special circumstances: When and only when every 'x' bound by '(Ex)' occurs within the scope of 'KaQa', in the case of (EKM), and 'PaQa', in the case of (EPM). In every other case whatever we automatically go to (EKM) or (EPM) and make use of distinct constant terms, a new term for each distinct modal profile.

Unfortunately, this set of qualifications is too strong and will have to be modified. Suppose we wish to instantiate upon '(Ex) (Fx & KaGx)'; by (EKM) we are allowed to infer 'Fc & KaGb', for some 'b' and 'c', but our world theory indicates that 'Fb & KaGb' ought to be deducible, for some 'b'. Therefore we must attach a qualification to our qualification. We shall say that in applying the rules (EKM) and (EPM) we may assign the same constant term to any pair of free 'x's' provided there is no occurrence of 'Pa' which binds one and not the other; in all other cases a new term must be assigned for each distinct modal profile. Similar remarks apply to the qualifications of (EK) and (EPKS) described above.

One final word about applying these rules. There may be uncertainty at times as to which rules a given formula falls under when it contains modal clauses sealed off from one another; for example, it is not at all clear in the case of '(Ex)(KaFx & PaGx)' whether or not (EK) may be employed. Therefore, we shall adopt the sort of convention described in footnote % and say that rules (EK), (EKM), and (UK) apply to any contexts in which 'x' occurs free within the scope of some 'Ka' operator itself not bound by any epistemic operator; only in cases where this condition is not met
will rules (EPL), (EPM), (EPS), (UPL), or (UPS) apply. Thus '(Ex)
(KaFx & PaGa)' does indeed fall under (EK).

***

In the last chapter we took notice of a "proof" offered by
Hintikka to show that Restricted Range is an inevitable consequence
of his model set approach to epistemic logic. Since we have now seen
that Hintikka's approach is perfectly compatible with giving up the
Restricted Range feature, let us diagnose Hintikka's argument and
see exactly where it breaks down.

The argument, we may recall, proceeded in roughly the following
manner. Suppose that all the rules of KB have been formulated except
for those governing formulas formed by quantifying in. Then (i)
Restricted Range is inevitable in KB if it is inevitable that '(Ex)
Ka(b=x)' is the weakest formula required to instantiate and
generalize upon formulas formed by quantifying in exactly one layer
of epistemic operators, to an arbitrary singular term 'b'. But (ii)
It is easily shown to be inevitable that '(Ex)Ka(b=x)' or something
which entails it is required to instantiate and generalize upon such
formulas. Therefore, (iii) Restricted Range is inevitable in KB.

What can be said of this argument? The following diagnosis is
offered by Sleigh in 'Epistemic Logic, Again':

A difficulty with this argument is the assumption that
\( L(108) \) and \( L(109) \) exhaust the conditions we might
wish to impose on instantiation and generalization
in simple epistemic contexts. This assumption seems to
me unfounded. Consider instantiation. Bearing in mind
that b ranges over definite singular terms including
definite descriptions it is clear that there are exactly
two distinct formulas containing b which have equal
right to be regarded as instantiations of '(x)Ka'
to b. 'Ka(y(b/x))' is the small-scope instantiation while '(Ex)(x=b & Ka(y))' is the large-scope instantiation. No doubt \( \mathcal{L}(109) \) yields the appropriate rules for small-scope instantiation. Without aiming for generality we might expect something like the following as a start toward covering the cases of large scope instantiation:

\[
(\forall x)Kax \in w \\
(\mathcal{E})(b=x) \in w \\
(\mathcal{E})(b=x & Ka(y)) \in w.
\]

With this condition (and no other changes made) we can now offer a proof that '(Ex)KaF' implies '(x)Fx'. This shows that Hintikka's argument for the necessity of restricted range is mistaken and suggests that the restricted range feature of KB is avoidable.16

Sleigh takes issue with premise (ii) of Hintikka's argument (in the form we have just restated it). It is true that '(Ex)Ka(b=x)' is required to generalize and instantiate upon the formulas in question, but it is true only with regard to one of two types of instantiation and generalization: small-scope instantiation. The much weaker condition '(Ex)(b=x)' suffices to do the job when we perform large-scope instantiation upon quantified formulas. Thus it is not in general true that '(Ex)Ka(b=x)' is required to generalize and instantiate upon formulas formed by quantifying in (where one epistemic operator in concerned); yet it is the general form of this claim upon which the argument depends. Hence the argument is mistaken.

The same criticism of Hintikka's argument seems to have occurred to Tienson. Consider:

(C.E)' If (Ex)Fx \in M, then Fb \in M and Qb \in M for at least one constant b

(C.U)' If (x)Fx \in M, and Qb \in M, then Fb \in M.

...we can take (Ex)Ka(b=x) as formulating the additional information we need for existential generalization and universal instantiation in the scope
of Ka and Pa, and we can substitute it for Qb in
(C.E)' and (C.U)'). This gives us $$\mathcal{L}(108)J$$ and $$\mathcal{L}(109)J$$.
It is clear that Hintikka takes this as establishing
that the basic restriction (Restricted Range) is
necessary. He does not make it clear why he does so.
Presumably it is because he takes this as establishing
that the rules $$\mathcal{L}(108)J$$ and $$\mathcal{L}(109)J$$ are necessary, and
in turn takes this as establishing the necessity of
the basic restriction. Such reasoning, however,
would be fallacious.....Although Hintikka constructs
a system in which the inferences in question hold
which contains the basic restriction, it is not
necessary that a system in which these inferences
hold contain the basic restriction as the argument
requires, for the addition of further rules concerning
quantification removes the restriction while leaving
the inferences intact.19

It seems to me that what Sleigh and Tienson have to say cannot
be disputed. Premise (i) seems true, but its truth depends upon
reading "instantiate and generalize" as "instantiate and generalize
both large and small-scope". Premise (ii) seems true, but it
truth depends upon reading the same phrase as "instantiate and
geneneralize small-scope". There is no reading according to which
both premises are true. Therefore the argument, at least in the
form we have stated it, seems to turn on an equivocation of the
phrase "instantiate and generalize". Hence either (i) or (ii) is
false, and the argument is unsound.

It should therefore come as no surprise that a number of systems
have been produced which are modelled after Hintikka's but which lack
the Restricted Range feature. It is inevitable in all of these systems
that '(Ex)Ka(b=x)' or something which is equivalent to it be required
for small-scope instantiation and generalization; this is precisely
what Hintikka's argument shows. But the argument shows nothing about
the inevitability of Restricted Range in epistemic logic.
Formula (c) is actually provable only with the rule \((C, \text{Inc.} =)\) which Hintikka did not add to his system until 1967.

As we have done in previous chapters, let us assume for the sake of simplicity that our concern throughout is with domains consisting of persons.

Føllesdal, 'Knowledge Identity, and Existence', p. 15.

Bearing in mind, of course, that the formulas in KB which are the formalizations of \((1')-(4')\) are not theorems in KB.

Tienson, 'The "Basic Restriction" in Hintikka's Quantified Epistemic Logic', pp. 5-6.

Sleigh, 'Epistemic Logic, Again', p. 6.

Ibid., p. 6. Hintikka himself discusses this possibility in Knowledge and Belief, p. 155, n. 12.

A word of explanation must accompany these rules. First, we assume there are no epistemic operators in \(F\). Second, we assume for purposes of simplicity that negation signs have been driven through epistemic operators. Third, we regard \((E)\) and \((U)\) as applying to any context in which \(x\) occurs within the scope of any \(Ka\) operator. The other rules apply to all contexts in which \(x\) is bound by \(P\) but not \(K\). In this way, \(\forall x(Fx & KaGx)\), \(\forall x(KaFx & KaGx)\), and \(\forall x(KaFx & PaGx)\) fall under \((E)\) while \(\forall x(Fx & PaGx)\) and \(\forall x(PaFx & PaGx)\) fall under rules other than \((E)\).

Sleigh, 'Restricted Range in Epistemic Logic', pp. 71-73. In doing so we shall in effect specify only the features of \(M\) which are relevant to the task of semantically evaluating our rules.

It might be felt desirable to regard \(I^*P_1\) as asserting that \(P_1\) is a rigid world-line. We leave ourselves uncommitted as to whether privileged world-lines in \(M\) are all and only the rigid world-lines in \(M\).

Hintikka later dispensed with this restriction.

\(W\) will not be a standard second-order theory in the sense that Existential Generalization on predicate constant terms is valid without restriction.

For proofs in \(W\) we use straightforward Copi-type rules with one modification: If \(\phi\) is of the form \(Fbw\), then from \(\phi\) may be
inferred 'B,bw', provided (i) 'B,' does not already appear in the proof, and (ii) The restriction "for some 'B,'" is attached to every formula in the proof containing 'B,'.

14 Tienson worries that without (*) there is no hope in proving such inferences as

\[
\frac{(x)Kafx}{-(Ex)Pa-Fx} \quad w.
\]

However, for us this is no worry since we have rules (C.P-) and (C.E-) which make all such inferences provable.

15 Perhaps it may be felt that no justification is needed to accept one set of English translations over another, and perhaps this is true. In any case, it is my wish to preserve a sort of isomorphism (which could be easily specified) between a formula's world theory transcription and its official English reading, at least for quantified formulas. If we accept the simpler set of English translations we accept them only in the capacity of non-official readings (readings obtained from official readings by equivalences). (5), (7), (8), and (9) remain the official readings and preserve the isomorphic relation between world theory and English.

16 Since the world theory transcription of '(x)(Kafx→Ka(x=x))' is a theorem in W, it better turn out that '(x)(Kafx→Ka(x=x))' is a theorem in KB*. It is.

17 Hintikka defines 'modal profile' in 'Existential and Uniqueness Presuppositions', p. 122.

18 Sleigh, 'Epistemic Logic, Again', pp. 5-6.

BIBLIOGRAPHY


Føllesdal, Dagfinn, 'Knowledge, Identity, and Existence', Theoria, 33 (1967), 1-27.


Hintikka, Jaakko, 'Modality and Quantification', Theoria, 27 (1961), 119-128.


———, "Knowing Oneself" and Other Problems in Epistemic Logic', Theoria, 32 (1966), 1-13.


167

---


---


---


---


---


---


---


---


---


---


---

''Knowing that One Knows'' Revisited', Synthese, 21 (1970), 141-162.

---


---


---


---


---


———, 'Epistemic Logic, Again', mimeographed, University of Massachusetts, 1972.


———, 'Hintikka on Quantification and Belief', *Nous*, 3 (1969), 399-408.

———, 'Two Women', *Philosophical Studies*, 23 (1972), 84-90.

———, 'Essentialism, Possible Worlds, and Propositional Attitudes', *Philosophical Review*, 82 (1973), 471-482.


———, 'The "Basic Restriction" in Hintikka's Quantified Epistemic Logic', mimeographed, Indiana University, 1972.