Faculty of Arts University of Helsinki

WORLDS AND OBJECTS OF EPISTEMIC SPACE

A STUDY OF JAAKKO HINTIKKA'S MODAL SEMANTICS

Tuukka Tanninen

DOCTORAL DISSERTATION

To be presented for public discussion with the permission of the Faculty of Arts of the University of Helsinki, in Hall 107, Athena Building, on the 18th of March, at 13 o'clock.

Helsinki 2022

ISBN 978-951-51-7883-1 (pbk.) ISBN 978-951-51-7884-8 (PDF)

Unigrafia Helsinki 2022

ABSTRACT

This study focuses on meaning and knowledge by assessing a distinctive view regarding their relation, namely the modal view of Jaakko Hintikka. The development of this view has not been previously scrutinized. By paying close attention to the texts of Hintikka, I show that, despite the extensive deployment of mathematical tools, the articulation of the view remained intuitive and vague. The study calls attention to several points at which Hintikka omits relevant details or disregards foundational questions. Attemps are made to articulate Hintikka's certain ideas in a more specific manner, and new problems that result are identified. The central claim argued for is that Hintikka's exposition was unsatisfactory in many respects and hence the view, as it stands, falls short in its explanatory scope compared to current theories in the intersection of logic, semantics, and epistemology. However, I argue that, despite its shortcomings, the prospects of the modal view are not exhausted. This is verified by introducing a new interpretation of the framework and by sketching new applications relevant in philosophy of language and in epistemology. It is also pointed out that certain early advances of the view closely resemble, and therefore anticipate, the central tenets of the currently influential two-dimensional approaches in logic and semantics.

ACKNOWLEDGEMENTS

I thank Gabriel Sandu for everything: not only did he hire me and guide me but he also gave me this topic, an office and a course to teach. I thank Vili Lähteenmäki for being a friend and for knowing everything there is to know about the academia. I thank my closest colleagues and friends Jaakko Hirvelä, Sanna Mattila, Pii Telakivi, Tuomas Vesterinen, and Maria Hämeen-Anttila for being there. I thank all the philosophy people in Helsinki and Tampere, especially Tuomo Aho, Hanne Appelqvist, Antti Kauppinen, Jaakko Kuorikoski, Simo Knuuttila, Tomi Kokkonen, Päivi Seppälä, Teemu and Pilvi Toppinen, Panu Raatikainen, Andreas Fjellstad, Fausto Barbero, Ilpo Hirvonen, Jani Hakkarainen, Markku Keinänen, Tommi Vehkavaara, Pasi Valtonen, Antti Keskinen, Anssi Korhonen, and Ville Paukkonen.

This project has received funding from the Finnish Cultural Foundation, the Eino Jutikkala Fund, and the Doctoral Programme in Philosophy, Arts and Society.

I thank the Finnish Doctoral Training Network in Philosophy for all the material and intellectual support.

I dedicate this work to Mira & Mesi, my bright lights here in the dark of the north.

CONTENTS

A:	BSTRA	CT	3			
A	CKNO	WLEDGEMENTS	4			
C	ONTE	VTS	6			
1	INT	TRODUCTION				
2	CRE	ATION OF EPISTEMIC SPACE	TION OF EPISTEMIC SPACE41			
	2.1	KRIPKE SEMANTICS BEFORE KRIPKE?	43			
	2.2	HISTORICAL STUDIES ON MODAL SEMANTICS	45			
	2.3	KRIPKE MODELS	45			
	2.4	HINTIKKA'S MODEL SETS	51			
	2.5	DEONTIC MODAL LOGIC	56			
	2.6	ALETHIC MODAL LOGIC	65			
	2.7	EPISTEMIC MODAL LOGIC	69			
	2.8	COMMENTS ON THE COMPARISON	75			
	2.9	QUANTIFYING INTO EPISTEMIC CONTEXTS	82			
	2.10	INCOMPLETENESS AND UNSOUNDNESS OF HINTIKKA'S LOGIC	86			
3	ОВЈ	ECTS	95			
	3.1	RIGID DESIGNATION	95			
	3.2	VARIETIES OF RIGIDITY	99			
	3.3	RIGIDITY IN MODEL SETS	103			
	3.4	WORLD-LINES AND NEO-KANTIANISM	104			
	3.5	GAME-THEORETICAL SEMANTICS	114			
	3.6	INDIVIDUALS AND OBJECTS	115			

4	WOI	RLDS	.117		
	4.1	PERSPECTIVAL CONTENT	. 119		
	4.2	TWO VARIANTS OF PERSPECTIVAL CONTENT	. 121		
	4.3	LIMITED ACCESSIBILITY	. 123		
	4.4	CENTERED CONTENTS	. 124		
	4.5	DIMENSIONS OF CONTENT	. 126		
	4.6	EMPTY SECONDARY CONTENT	. 139		
	4.7	HINTIKKA-STYLE COMMUNICATION	. 139		
	4.8	METAPHYSICALLY OR EPISTEMICALLY POSSIBLE WORLDS?	. 153		
5	CON	NCLUSION	. 155		
Ri	References				

Introduction

1 INTRODUCTION

We know that Jupiter has several moons. Galileo knew that too. Galileo spotted the moons through his telescope and sketched a series of illustrations depicting their relative positions. In the received philosophical jargon we share the same propositional attitude with Galileo. The proposition is what Galileo's drawings represent and what the that-clause in the first sentence of this paragraph means or expresses, namely that Jupiter has several moons. Our shared attitude towards that proposition is knowing.

This study discusses the key notions of the story above by assessing a distinctive view regarding propositional attitudes, namely the modal view envisaged by G. H. von Wright and developed in detail by Jaakko Hintikka. The development of this view has not previously been scrutinized. The main findings, from a historical point of view, include, first, that the view came in three radically different variants because its development took place in three stages, each driven by distinct methodological considerations. Second, each variant has its own specific shortcomings and hence the view, as it stands, faces challenges with respect to its explanatory scope compared to some of the current theories in the intersection of logic, semantics and epistemology. To understand why this is so, it is necessary to pay close attention to the texts of Hintikka with a special focus on the technical notions of *modal space*, *individual*, and *possible world*. It is shown that, despite the extensive deployment of mathematical tools, the development of the view remained intuitive and vague. The study calls attention to several points at which Hintikka omits relevant details or disregards foundational questions.

From a more systematic point of view, attemps are made to articulate some of Hintikka's ideas in a more specific manner, and new problems that result are identified. I do not claim that these specifications are the ones that Hintikka would have had pursued or that they are the only possible ones. But I argue that the specifications studied here are plausible developments of the ideas under scrutiny.

More specifically, Chapter 2 shows that Hintikka's first semantic framework was built on syntactic foundations and thereby certain semantic questions fall outside its scope. Further, it is argued that some of the obvious attempts to enrich the framework with sufficient semantic details lead to intolerable problems. Chapter 3 argues that Hintikka's later attempts to solve these problems with entirely new conceptual tools were afflicted by incoherent ideas concerning individuals. An attempt is made to re-articulate Hintikka's view of semantic neo-Kantianism in order to provide a more coherent outlook. Chapter 4 deploys this new outlook to provide a new interpretation of the framework and address some of the questions that the currently influential two-dimensional approaches in logic and semantics also seek to address. I also emphasize and critically assess certain

early advances of the view that closely resemble and thereby anticipate the central tenets of two-dimensional semantics.

What, then, is this modal view of Hintikka? It is fundamentally based on the consideration of possibilia such as alternative states of affairs, alternate histories, or more commonly, possible worlds. With this talk of "possible worlds" the view may strike one as literally otherworldly. However, it may be understood as deriving from a simple and commonsensical picture of information and (scientific) investigation. According to this picture, all information about the world is information about which possible world is the actual one, that is, information about the location of our world in the space of possible worlds. Roughly, all our information about the world may be identified with the set of possible worlds we cannot rule out on the basis of empirical evidence and (ideal) reasoning. As investigation provides us more evidence, we may narrow down the range of possibilities, as Galileo was able to do during the first nights of the year 1610. On January 6th Galileo had no evidence of Jupiter's moons. The possibilities compatible with Galileo's evidence included some possible worlds with no moons at all and some worlds with several hundreds of moons in the orbit of Jupiter. The first observations provided sweeping evidence as Galileo spotted three moons. Then he was able to narrow down the possibilities by excluding all worlds with two or less moons and to establish the revolutionary result that Jupiter has at least three moons. The following observations brought a fourth moon into his view, and this made the result more precise since he was able to exclude all worlds with three or less moons.

Since one of the main purposes of language is to transmit information about the world, this picture combines naturally with an account of semantic content according to which the content of a sentence is the information encoded in it. Since any piece of information may be identified as a set of possible worlds we may also think of any semantic content of a meaningful sentence as a set of possible worlds.

In general, the modal view analyzes a propositional attitude as a relation between a subject and a set of possible worlds. In Galileo's case, the worlds are those in which the sentence *Jupiter has several moons* is true (and Galileo's drawings capture the state of affairs correctly). In a sense, that set of worlds *is* the proposition, which serves, first, as the content (or at least a central aspect of the content) of the sentence *Jupiter has several moons* and, second, as the object of our knowledge and other attitudes.

This picture connects modal, epistemic, and semantic considerations in an elegant way, but it has come under considerable strain due to the widely accepted separation of epistemic and metaphysical modality, as we shall shortly see. Frank Jackson outlines the general modal picture connecting information, communication, and truth as follows:

Consider what happens when I utter the sentence 'There is a land-mine two meters away.' I tell you something about how things are, and to do that is precisely to tell you which of the various possibilities concerning how things are is actual. My success in conveying this urgent bit of information depends on two things: your understanding the sentence, and your taking the sentence to be true. We have here a folk theory that ties together understanding, truth, and information about possibilities; and the obvious way to articulate this folk theory is to identify, or at least essentially connect, understanding a sentence with knowing the conditions under which it is true; that is knowing the possible worlds in which it is true and the possible worlds in which it is false; that is knowing the proposition it expresses on *one* use of the term 'proposition'. This kind of theory in its philosophically sophisticated articulations is best known through work of David Lewis and Robert Stalnaker. (Jackson 1998a, 71)

Jackson had it right: the view is probably best known through the writings of Lewis and Stalnaker. At least, some of the central doctrines of this view are commonly associated with these writers, namely that the semantic contents of sentences, the objects of thought, and the primary bearers of truth and falsity (in short: propositions) are sets of possible worlds, and that to gain knowledge and to specify it is to narrow down this set. However, the first "philosophically sophisticated articulations" of this view are not the ones that are best known. The first articulations were sketched by von Wright in the early 40's and published in his *An Essay in Modal Logic* (1951a), and subsequently the approach was developed in detail with many variants of possible worlds semantics by Hintikka in his 1957 papers and especially in his seminal book *Knowledge and Belief: An Introduction into the Logic of the Two Notions* (henceforth *K&B*), published in 1962.

In the monographs von Wright and Hintikka introduced a novel orientation: "The modal categories we shall understand in a somewhat wider sense than is usually the case", von Wright (1951a, 3) wrote. Von Wright was among the first to systematically utilize an idea considered foundational in contemporary thinking about modality, namely the idea that the modal notions are logically reducible to quantifiers: the seemingly peculiar inferential properties of modal notions are, in fact, just the familiar inferential properties of quantifiers. Further, von Wright saw that obligation, knowledge, belief, and many other notions are conceivable as

¹ The previous picture of information is simply extended to knowledge. If an epistemic subject has no knowledge then all epistemic possibilities are open to her. If the subject knows that p then all the possibilities in which p is false are ruled out. If the subject knows that that p and q then all the possibilities in which p and q are false are ruled out, and so forth. If the subject is omniscient then no epistemic possibilities are left open save one, the actuality.

² Generally these ideas have been around since antiquity, that is, modal concepts have been regarded as similar to the concepts of "some" and "all" and modal statements have been regarded as involving the evaluation of sets of possibilities of some kind.

modal notions on par with alethic modalities of necessity and possibility in the sense that similar logical systems can be employed to study their structure.³ As the result, the alethic modal space invoked by the early possible world semanticists such as Carnap (1947) was transformed and gradually developed by Hintikka into an epistemic space where worlds may be logically impossible in the sense that contradictory sentences may truly describe them, and the objects occupying them may resist customary criteria for identity such as the variants of the principle known as the *Leibniz Law*. This raises a spectrum of questions ranging from interpretational to ontological. The first set of answers, or rather the first set of commitments, was introduced along Hintikka's first meta-linguistic framework.⁴ In Chapter 2 this first phase is studied in relation to the emergence of possible worlds semantics and some of the traditional puzzle scenarios and related problems concerning propositional attitudes.

The second phase was marked by Hintikka's rejection of the syntactically driven model set semantics in the transitional chapters of *Models for Modalities* (1969), such as "Semantics for Propositional Attitudes", which introduced proper model-theoretic semantics with individuating functions that track objects across modal space. Later these functions became known as *world-lines* and hence the approach may be dubbed *world-line semantics*. Hintikka stressed that we must "interpret 'possible worlds' realistically" (Hintikka 1975, 96) as "real-life alternatives to our actual world" (Hintikka 1975, 84) and the values of the variables of our formal language must be "real, fullfledged individuals" (Hintikka 1967a, 38) while we must simultaneously take into account "impossible possible worlds" and "manifestations" or "stages" of individuals from different subjective perspectives. The latter are certainly not "real, fullfledged individuals" but something entirely different, perhaps mere echoes or shadows cast by real individuals across epistemic space. Chapter 3 studies the nature of Hintikka's individuals.

In a wider picture, Hintikka concluded that the alethic space (how things are) and the epistemic space (how things appear to a subject) must be first distinguished in some sense and then combined into a single framework in one way or another. Numerous studies under the rubric "two-dimensional semantics" have recently been published evaluating the prospects and the plausibility of this very idea.

The backdrop of two-dimensional approaches was Kripke's lectures of 1970 and their aftermath, which convinced philosophers that there are necessary *a posteriori* truths (as well as contingent *a priori* truths). An important implication of *a posteriori* necessities in the context at hand is that the alethic or metaphysical

³ Also C. S. Peirce distinguished an epistemic concept of possibility from what he called "substantive possibility" (alethic possibility), and regarded modalities as quantifiers over "possible states of things" (Peirce 1931–35: vol. II, paragraph 2.347; vol. V, paragraphs 5.454–455). Peirce regarded epistemic concepts as modal concepts, but he did not develop epistemic logic in a systematic way.

⁴ Apparently Hintikka received considerable assistance from von Wright in this project even though Hintikka mainly took care of the technical development and publishing, see e.g. Hintikka (1957b, 2). Von Wright explicitly deploys this framework only in *Norm and Action* (1963, 19–22).

modality and the epistemic modality get separated. They may overlap, but in order to accommodate the *a posteriori* necessities one must assume that some epistemic possibilities are not metaphysical possibilities. In terms of possible worlds, a necessary truth is true in all metaphysically possible worlds and an *a priori* truth or *epistemically necessary* truth is true in all epistemically possible worlds. If these sets of worlds are the same then, *prima facie*, the necessary and the *a priori* go hand-in-hand: what is necessary is always *a priori* (and what is contingent is *a posteriori*). If a metaphysically necessary truth *p* is to be *a posteriori* then there must be room in the analysis for epistemically possible worlds in which *p* is false (that is, worlds that may be ruled out by empirical inquiry). Hence alethic and epistemic worlds, or rather modal spaces, must come apart.

But these considerations blur the elegant picture of information, inquiry, and content. Should we identify information and linguistic content with sets of metaphysically possible worlds or epistemically possible worlds, or a combination of the two? All three options come with serious problems. Intuitively, epistemic possibility and metaphysical possibility are very different: the first has to do with what we may conceive or imagine or conceptualize while the latter has to do with how things could be. If we choose to identify information and content with the sets of metaphysically possible worlds, then we must, in order to accommodate Kripke's examples, give an account of the epistemic dimension involved in terms of metaphysically possible worlds. If, on the other hand, we confine ourselves to sets of epistemically possible worlds then we must give an account of metaphysical dimension in terms of epistemically possible worlds. Both positions are radical, potentially involving a category mistake of confusing conceptual and physical realms. Things are not easy for a modal dualist either: she must address the question of the relationship of the two distinct modal spaces and their interaction in the matter at hand.

Some variants of two-dimensional semantics have been put forth to re-unify these modalities in the monistic spirit (e.g. in Jackson 1998a; 2011). Hintikka entertained similar ideas in the mid 60s and implemented them in his formal semantics in the chapter "On the Logic of Perception" of Hintikka (1969). Hintikka did not explicitly consider the alethic dimension as metaphysical possibility but rather as the objective physical reality that scientific inquiry is concerned with. However, I argue that Hintikka anticipated the relevance of different dimensions in modal semantics and pioneered their formal implementation, and this fact has been missed in the literature. For instance, the extensive editions by Lepore & Smith (2006), Garcia-Carpintero & Macia (2006), and Davies & Stoljar (2004), which each includes long sections or whole papers on the history of two-dimensional approaches in modal semantics, contain no reference to Hintikka's contributions. These contributions are discussed in Chapter 4, which also invokes the difficult question concerning the exact nature of possible worlds. This topic is not properly discussed here but certain conclusions

based on Hintikka's ideas are suggested. One of the very few things Hintikka (1969, 171; 1975, 195) said about this question was that he is concerned not with complete worlds but "small worlds", which are understood as quite limited and local scenarios about, for instance, how a person could have perceived things differently in a given situation.

In general, both frameworks at hand are surprisingly poorly known. Many consider K&B as a classic text of analytic philosophy, or at least it is frequently referred to in the contemporary literature as a source several seminal ideas. The precise content of the book, however, has received little attention. The main reason for this is that the formal semantics for K&B's logic, model set semantics, has been poorly understood.⁵ The subsequent world-line framework is also poorly known but for different reasons, one being that Hintikka never fully developed it in print. For instance, Hintikka's most important single idea concerning the union of the alethic and epistemic spaces, namely the distinction between the physical or public method and the perspectival method for individuating objects (accompanied by the parallel formal distinction between the two modes of quantification), has remained obscure in its details despite the fact that Hintikka wrote dozens of papers about it (including one of the very last he published in his lifetime, "Epistemologia vetus et nova", in 2013). Several symptomatic quotes can be found illustrating Hintikka's level of generality and his unwillingness to sort out the details of his distinctions. Here is one:

How we cross-identify between these worlds by those usual public or descriptive methods which we are apt to think of first in this sort of context is both so obvious in its general features and so problematic in its details that there is little point in discussing it here at great length. (Hintikka 1975, 64)

Consequently, the distinction (and the world-line framework with it) has been neglected in the literature. For instance, only one author, Risto Hilpinen (2006), assessed it in the *Library of Living Philosophers* edition dedicated to Hintikka's work. Apparently Hintikka was satisfied neither with the omission nor with Hilpinen's exposition. Hintikka (2006, 32, 823) commented that the "insight into the duality of our actual identification methods is [...] the most important neglected idea in contemporary analytic philosophy" and regarding Hilpinen's contribution he wrote that the "implications of these parallel distinctions are so sweeping that no one can exhaust them in one article".

⁵ The situation is made worse by the fact that the only edition of $K\mathfrak{C}^{\bullet}B$ available in print, namely the Hendricks & Symons edition of 2005, contains some editorial blunders: for instance, section 6.14 suggests that Hintikka proved by using model sets that formulae obtained from the schema $K_a \nabla_x Fx \to \nabla_x K_a Fx$ (which is an epistemic counterpart of the converse Barcan formula) are both valid and not valid.

Another reason why Hintikka's second framework is relatively unknown is that it has been considered a contribution to logic, which it is not, or at least not exclusively. Various commentators throughout the years, for instance Føllesdal (1969), Stalnaker (1972), Lycan (1994), and Perry (2009), have considered Hintikka's second framework as an (ill-conceived) attempt to provide foundations for quantified epistemic logic. According to Stalnaker (1972, 458) Hintikka's approach was too "Protagorean", that is, too relativistic to justify any logic while Perry (2009, Section 6) evaluated that it as simply too "odd". I claim that the motivation for the second framework lay elsewhere, namely in combining the alethic and epistemic spaces into a program with much broader philosophical goals than mere formal semantics for quantified formulae. Hintikka's "Protagoreanism" and "oddness" should be evaluated in an appropriate context: his approach should be compared with the variants of two-dimensional semantic frameworks or with situation semantics by Barwise & Perry (1983), or Lycan's (1994) mature linguistic Ersatzism. It is interesting that many reviewers who criticized Hintikka's second framework in the context of epistemic logic (e.g. Stalnaker, Perry, Lycan) have constructed general epistemo-linguistic frameworks of their own reflecting many of Hintikka's early ideas - without giving credit to Hintikka.

Due to books such as Reasoning About Knowledge by Fagin et al. (1995) the propositional epistemic logic based on Hintikka's ideas has become not only wellknown but also a standard tool in computer science, with a wide range of applications in fields such as economics, artificial intelligence, and game theory. However, propositional epistemic logic was merely a starting point for Hintikka's program. The goal was to construct a general framework backed up by rich formal resources far beyond the expression power of propositional languages to reveal not only the so-called "formal properties" of epistemic attitudes but to clarify and model the use of epistemic expressions in natural language and depict the nature of epistemic representation and misrepresentation in general, all the way to the explication of the structures of thought and thought contents of knowers and believers while taking into account the empirical data provided by the latest neuroscience (see e.g. Hintikka 1990; Hintikka & Symons 2003). These are not goals of epistemic logic but goals of a much broader philosophical project. The goal of (static) epistemic logic is to model the information that agents have and how they reason with it, and this was indeed one of Hintikka's early goals. But his scope had already widened significantly in K&B, or so I argue.⁶

To recap, after the appraisal of Hintikka's first model set framework, I assess the second with world-lines and show how some of the problems that twodimensional semantics was designed to solve may also be addressed by the means

⁶ It is evident that Hintikka did not introduce e.g. the so-called impossible possible worlds for logical purposes but for semantical and philosophical purposes. Inconsistent model sets ("impossible worlds") corrupt the proof procedure of *K&B* and other attitude logics and the notion of validity provided by the model set technique. Impossible worlds were never meant to be a part of Hintikka's logic.

of a suitably specified Hintikka-style approach. I present a novel application in communication theory built on the aforementioned distinction between the physical and the perspectival mode of individuation. Other central notions or topics in this study include puzzles, intuitions, reference, intentionality, rigidity, propositions, the ontology/epistemology interface, the internalism/externalism divide, de re attitudes, and dimensions in modal semantics, among others. Below I offer some introductory remarks concerning these notions that comprise the philosophical background of the study. Recognition of the surrounding conceptual landscape is crucial to the purposes at hand: Hintikka's work cannot be evaluated without regard for this context. But first, let me point out two main omissions in this study.

The main historical oversight is that Hintikka's extensive studies relating to Kant are largely suppressed, despite their undeniable relevance to topics at hand. Many of Hintikka's writings on Kant (for instance the earlier ones collected in Hintikka 1973) are specifically about meaning and knowledge, that is, about the main topics of this study. However, the main focus of Hintikka's Kant interpretations and variations was on the philosophy of mathematics and thereby the discussions of meaning and knowledge emphasized the aspects of *analyticity* and *a prioricity* in the context of mathematical reasoning. These notions are not among the main concerns here. Second, there exists no proper general exposition of Hintikka's writings on Kant. Such an exposition would, of course, be most welcome but a pioneering general assessment of the intricacies of that overwhelming material is a matter of another independent study.

On the more systematic side, the problem of logical omniscience is not discussed here in detail. Omniscience is a major problem in epistemic logic by itself, but it also creates an interesting double bind when combined with the socalled Kaplan's paradox (Kaplan 1995). This joint-problem undermines in particular all the approaches that apply possible worlds in the analysis of propositional attitudes. One set of problems arises when one's possible world semantics does not distinguish between logically equivalent expressions: one has two expressions that intuitively differ in meaning but determine the same set of worlds and hence, according to one's semantic theory, express the same proposition. In the setting at hand, this leads to the conclusion that if one bears an epistemic relation to a mathematical or logical truth then one bears an epistemic relation to all mathematical and logical truths, since these are necessarily equivalent by virtue of determining the same set of worlds. The standard move is to introduce a more fine-grained conception of propositions. A more fine-grained view may be sensitive, for instance, to the structures of propositions and may treat Bachelors are unmarried and Brothers are male siblings as expressing distinct propositions due to their different structures (assuming that they express logical truths and are thereby equivalent).

⁷ Sintonen (1994) provides some guidelines for such a project.

Kaplan's paradox, on the other hand, has it that if the number of the elements of the set of all possible worlds has the cardinality K, then according to all views that identify propositions with sets of possible worlds, the number of propositions, that is, the number of the subsets of K, must be 2^K . But if we assume that for each proposition p and each time t_0 there is a possible world in which Kaplan entertains precisely p and no other proposition at t_0 , then there is a bijection from the set of propositions to the set of possible worlds: that is, there is a one-to-one mapping of the elements of a power set of a set to the elements of the set itself. But this contradicts Cantor's power set theorem. Further, a more fine-grained notion of propositions only makes this problem worse! Hence there are interesting connections between these problems. However, neither these connections nor the separate problems are studied here. Two things may be said in defense of such an omission. First, distinguishing necessarily equivalent expressions is not exclusively a problem for possible world treatments of propositional attitudes - it is a problem for all semantic theories. Second, strictly taken, the historical remarks and the systematic applications introduced here do not concern the set of all possible worlds, let alone its power set. Nevertheless, I do not underestimate the relevance of these problems. On the contrary, they deserve attention, but they cannot receive that attention here.

Hence, concerning both omissions let me re-state that this study is mainly about the problems that arise in the case of everyday contingent synthetic *a posteriori* propositions and attitudes, such as *I know that there's fish for dinner*. An analysis of such an attitude with the tools at hand may be interesting and useful without an articulated position concerning the epistemology of mathematics or paradoxes of Set Theory. Next, I shall comment on the central notions of this study.

Puzzles and Intuitions

Picture us, you and me, standing in an open landscape, gazing into the distance. We see a man approaching. Due to my poor eyesight, I do not recognize him but you, on the other hand, have no difficulty in identifying him. Let us say that the man is Socrates. Thus we see Socrates approaching, or at least you do. But do I see who is approaching? According to 14th century scholars, the answer may depend on whether the question is considered in sensu diviso (in the divided sense) or in sensu composito (in the compound sense). For instance, William Heytesbury might have held that the answer is positive if the main verb see and the grammatical object Socrates are divided as the underlining indicates in the conclusion below:

That I see approaching,
That is Socrates;
Therefore, <u>Socrates</u> I <u>see</u> approaching

In the compound sense, however, the answer is negative: while the above syllogism is valid, the following is not:

That I see approaching,
That is Socrates;
Therefore, I see Socrates approaching⁸

Such puzzles, scenarios or thought experiments, have been paradigmatic in the literature on propositional attitudes. Contemporary formulations are usually variations of Frege's (1892) and Russell's (1905) famous examples.

Some philosophers have a low opinion of puzzles. Hintikka was one of them (despite the fact that he studied puzzles involving the notion of knowing who, as we shall see). In his "Methodological Introduction" (Hintikka & Kulas 1985) he complains that it is misguided for a theorist to rely on some intuition about this or that puzzle scenario invented by the theorist herself, and then to generalize the data and the moral of the puzzle to serve as a basis of linguistic theorizing. Hintikka writes that one's intuitions about a given puzzle scenario are not necessarily shared by others; intuitions are subjective and simply not reliable and sharp enough. According to Hintikka (1999, 137), contemporary analytic philosophers have acquired an intuitionistic methodology committed to an illusion called the "atomistic postulate", according to which "the basic input into our epistemic process consists of particular data, excluding general truths". According to Hintikka, philosophical views should be systematically developed in the context of general frameworks such as model theory and modal logic as Hintikka himself did, and not in the context of intuitions invoked by descriptions of singular scenarios. For these reasons, Hintikka writes, he has "tried to rely more on quotations from contemporary fiction, newspapers and magazines than on linguists' and philosophers' ad hoc examples" as sources of data (Hintikka & Kulas 1985, ix). Indeed, Hintikka wrote very little about the famous puzzles, and if he did, he usually did so in a ridiculing tone. For instance, Hintikka solved "Kripke's home-made puzzle" about belief in one paragraph by positing a function that took care of the problem in his formal semantics by fiat - without offering any explanation of what actual aspect of language or its use the function represented, except that the function did not represent "ascribing descriptive content to the two proper names" (Hintikka & Sandu 1995, 278). The message is clear: Kripke's puzzle is an ad hoc problem and what it deserves, at most, is an ad hoc solution.

Some take puzzles more seriously. Kripke wrote about the so-called *surprise* execution puzzle as follows:⁹

⁸ Similar examples can be found in Heytesbury's *De scire et dubitare* in Kretzmann & Stump (1988, 444, passim).

⁹ A prisoner is sentenced to death by a surprise execution that takes place within five days. The prisoner reasons that the execution cannot be on the fifth day because then it would not be a surprise. But then the execution cannot be on the fourth day either for the same reason; and it cannot be on the third day, and so on.

It is interesting that this kind of problem [the surprise execution puzzle] is discussed as if it were a philosophical problem at all. How philosophical it actually is depends on whatever philosophical morals we may draw from it. Graham Greene classifies his works into novels, entertainments, and some other works: a novel is supposed to be a more serious work, but entertainments are often the best. A problem like this might be classified as an entertainment in this sense. But it can have aspects of a "novel" if conclusions concerning our basic concepts of knowledge may be drawn from it. Here, more so than with typical philosophical problems, we are in the kind of "intellectual cramp" that Wittgenstein describes — one in which all the facts seem to be before us, there does not seem to be any new information to be gained, and yet we don't quite know what is going wrong with our picture of the problem. (Kripke 2013, 28)

Puzzles are ways to frame philosophical problems, among many other ways. A newspaper headline, a mathematical proof, or a sublime passage from *The Republic* may frame a philosophical problem. If this or that framing allows us to draw interesting philosophical conclusions it should not matter which framing prompted them. If an interesting conclusion concerning our basic notions may be drawn, who cares where the data came from? But perhaps Hintikka's point was that traditional philosophical puzzles have generated poor discussions and poor, merely local conclusions.

One important feature of philosophical puzzles, at least from the viewpoint of the present study, is that at their best they force one to pay attention to the smallest details of one's framework. Hintikka's writings concerning his second framework would have benefited greatly if he had paid closer attention to the details of, say, Frege's puzzles, since they capsulate some of the central questions that Hintikka's world-line framework was designed to answer: why does the identity of Hesperus and Phosphorus seem to be *neither* an objective alethic fact independent of our conceptualizations *nor* merely an epistemic or semantic fact constituted by our conceptualizations? If it is a fact of both natures, how exactly do the alethic and epistemic notions interplay in the case of identity statements? Questions such as these led Hintikka to distance himself from his early model set semantics, but he never applied his new apparatus to the puzzles of identity properly in print, and hence he was never forced to explicate some of the central ideas of his frameworks.¹⁰

¹⁰ Hintikka commented on Frege's identity and propositional attitude puzzles in many of his writings, e.g. in Hintikka (1957a), KℰB, and Hintikka (1980a), but each time he basically repeated the general remarks he had already made in Hintikka (1957a): the puzzles arise because the involved singular terms designate different individuals in different possible worlds.

There are also puzzles that are not *ad hoc* or "home-made". Consider the following: $\forall x(x = x)$ is the law of universal self-identity from classical predicate logic with identity, and hence something that a cognitively ideal agent of Hintikka's epistemic logic knows right at the outset. Hence we may write

$$K_{a} \forall x(x = x)$$

that is, a knows that $\forall x(x = x)$. This is, of course, problematic in its own right: a Hintikkian logically omniscient agent knows every valid formula. But it gets worse. Since Hintikka (1962, 163) proves the validity of the epistemic counterpart of the converse Barcan formula in the logic of $K \not \subset B$, we may derive a stronger statement, $\forall x K_a(x = x)$, that is, of every object a knows that it is self-identical.

- (i) $\forall x(x = x)$
- (ii) $K_a \forall x (x = x)$
- (iii) $K_a \forall x (x = x) \rightarrow \forall x K_a (x = x)$
- (iv) $\forall x K_{\alpha}(x = x)$

(i) is the aforementioned logical law: (ii) follows by the epistemic counterpart of necessitation, or the *knowledge generalization rule* as it is sometimes called; (iii) is the relevant epistemic instance of the converse Barcan formula, a theorem of the $K \not\sim B$ logic; and (iv) follows by *modus ponens*. Now, necessitation and the converse Barcan formula are both problematic in the context of epistemic reasoning. However, we may also prove $\forall x K_a(x = x)$ using Hintikka's model set rules without appealing directly to these principles.¹¹ If so, then we may without any provisos (that for instance Kripke (1971) has to discuss in order to justify $\forall x \square (x = x)$ as a postulate) continue with an epistemic counterpart of the standard derivation of necessity of identity via indiscernibility of identicals (also known as the *Leibniz Law*):

(II)
$$\forall x \forall y ((x = y) \rightarrow (Fx \rightarrow Fy))$$

and conclude that for all x and y, if they are identical then they are known to be identical:

¹¹ Hintikka did not provide a proof of $\forall x K_a(x = x)$ but it is easy to construct. For a Hintikka-style *reductio* proof, suppose that there is model set μ such that it contains the negation of $\forall x K_a(x = x)$:

^{1.} $\neg \forall x K_a(x = x) \in \mu$

^{3.} $\exists x K_a(x = b)$ $\in \mu$

^{5.} $\neg (b = b)$ $\in \mu^*$ from (4) by (C.P*)

⁽⁵⁾ is a contradiction that violates Hintikka's condition (C.self \neq). Hence (1) is not satisfiable and thus the formula $\forall x K_a(x = x)$ is valid. (Hintikka's proof procedure and model set rules are discussed in Chapter 2 below).

- (iv) $\forall x K_{\alpha}(x = x)$
- (v) $\forall x \forall y (x = y) \rightarrow (K_a(x = x) \rightarrow K_a(x = y))$
- (vi) $\forall x \forall y ((x = y) \rightarrow K_{\alpha}(x = y))$

(iv) is valid as shown; (v) is the relevant instance of (II) in which the property attributed to both x and y is the modal property of is known to be identical with x, and (vi) follows by classical predicate logic.

Now, the puzzle is: what should we think about (vi)? Should we consider it as valid? Are all identities known identities? If not, what is wrong with (i) - (vi)? Hintikka said very little about this question and it remains open in relation to model sets as well as to world-line semantics.

Years after K&B, when Hintikka was distancing his work from the model sets he returned to this question and introduced, in the transitional chapter "Existential Presuppositions and Uniqueness Presuppositions" of *Models for Modalities* (1969), the notion of *modal profile* together with the highly complex model set conditions (C.ind=), (C.ind=0), and (C.ind=E) that were designed, once again, to guarantee a rigid interpretation of variables (a recurring theme in Hintikka's work, as we shall see below). He tried to prove (II) and claimed that, by the means of his new conditions, the necessity of identity may also be proved:

(NI)
$$\forall x \forall y ((x = y) \rightarrow \Box (x = y))$$

The discussion was in the context of alethic modal logic. Concerning the validity of the epistemic counterpart of (NI), that is, (vi), in epistemic logic, Hintikka merely mentioned that

There are rather plausible-looking counter-examples [...] in epistemic logic [to (vi)]. I believe that I can nevertheless explain them away. To attempt to do it here would take us too far, however. (Hintikka 1969, 130)

Hence the question remains open. It should be noted that, first, this puzzle arises when one is trying to make sense of quantifying into epistemic contexts in accordance with the usual logical principles of quantification and identity. It does not appeal to vague intuitions, only to the fact that some identities are unknown, or more generally, to the fact that truth does not entail knowledge. Second, this puzzle concerns variables, the most basic artifice of our modeling, and the objects assigned as the values of variables. Accounts of proper names, descriptions, indexicals or other natural language expressions are irrelevant. We may take a descriptivist or a direct referentialist (or whatever) position concerning the semantic import of the aforementioned expressions, but that does not affect the

central difficulty: if we regard quantifying into epistemic contexts as meaningful, we must address this puzzle on the level of variables. However, Hintikka left this puzzle unanswered. The end of Chapter 2, especially section 2.10 below, assesses some of the consequences of this neglect.

Intention and Reference

Hintikka's approach to propositional attitudes may be characterized by the slogan *intentionality first*. In $K \mathcal{C}B$, Hintikka seemed to take the notion of belief to be the most central in his logico-epistemic theorization, which of course is the traditional stance. Intuitively, forming beliefs is the basic function of the human mind, since all epistemic attitudes seem to be some sort of beliefs. To know, to perceive, to recall, to hope, or to want is to believe, that is, to take something to be the case (under some qualifying conditions).

One of Hintikka's most important post- $K \otimes B$ ideas was that the logical analysis of knowledge and belief is to be extended to other attitudes. This possibility opened up when Hintikka took intentionality to be the basic concept. All the aforementioned attitudes, including belief, may be understood as intentional attitudes, and the analyses of other attitudes from hallucination to memory follow by setting further conditions to intentionality.

Hintikka came up with the idea alluded to in the title of his 1975 book The Intentions of Intentionality and Other New Models for Modalities, namely the idea that intensionality with an s is the distinguishing characteristic of intentionality with a t in the sense that an intentional concept is such that its semantics requires considerations of possibilia, that is, relational models provided by possible worlds semantics (Hintikka 1975, 194-195). This idea seems dubious today. Intentionality is defined in the relevant Stanford Encyclopedia entry as "the power of minds to be about, to represent, or to stand for, things, properties and states of affairs". Metaphysical necessity/possibility requires considerations of possibilia, but very few, if any, would consider it as an intentional concept in any even remotely Husserlian sense. Metaphysical necessity/possibility is not a propositional attitude; it is by a textbook definition a non-epistemic notion and hence it does not relate to the "powers of mind". Further, the idea that intensionality is the hallmark of intentionality must have seemed dubious even in the early seventies since, even though Hintikka and many of his collaborators were perhaps not very well aware of the concept of metaphysical necessity/possibility (or if they were, they could not have foreseen how influential that concept would become) they certainly were aware of the related concept of physical possibility in the sense that one may ask what is and what is not possible according to the actual laws of nature. For instance, I may ask whether it is physically possible for me to stand on Jupiter, and the answer requires considering physically possible worlds in which the actual laws of nature hold, in order to characterize the constraints under which it may be said to be possible or impossible for me to stand on Jupiter. The answer obviously

requires considerations of *possibilia* but the notion of physical possibility is not intentional.

Nevertheless, the importance of the notion of intentionality in Hintikka's thought cannot be overestimated, despite the fact that Hintikka did not pursue a more detailed definition of intentionality. Hintikka (1975, 17) referred to Chisholm's work in this area and stated that he aimed to provide "sharper tools" for approaching Chisholm's ideas. Indeed, many of the Chisholm's views make perfect sense in Hintikka's modal environment. Both Hintikka's intentionality first approach and Chisholm's (1984, 89) principle of "primacy of the intentional" take the intentionality of thought to more basic idea than reference of language: linguistic expressions refer to objects only in so far as they are used express thoughts about objects. This contrasts sharply with many current methodologies in philosophy of language that emphasize the intrinsic referential nature of certain expressions or the foundational presence of objects in a certain class of propositions and seek to explain our object-directed thoughts in terms of these referential expressions and/or the so-called singular propositions. Chisholm and Hintikka start with an intentional thought while the nature of reference and propositions are not presupposed but questions we must investigate and draw, perhaps reluctantly, conclusions about. For both, individuation is the road to reference. 12 In order to have an object-directed thought one has to single out the object, for instance, in one's visual field. In order to refer to an object, to pick it out among other things, or to state something about it, one must individuate it first. Both Chisholm (1984, 97) and Hintikka (1969) consider perceptual reference, a topic that most theories of reference do not deal with, as the most basic reference. This emphasis is one aspect in Hintikka's thought that makes the notion of rigidity problematic. We perceive properties, not objects themselves. Perceptual reference operates via properties while an object is merely a property cluster from a perceptual point of view. Hence the objects of perceptual attitudes are not "propositions" but first and foremost properties and derivatively objects individuated via properties. What does it then mean to ask of a cluster whether or not it is the "same" object in every possible world? I assess this question in Chapter 3.

After $K \mathcal{C}B$, Hintikka acquired a more substantial focus to reference, which as a phenomenon connects our language and, more generally, our epistemic attitudes to the world, and hence sets the stage for the discussions of intentionality in the context at hand. In $K \mathcal{C}B$ reference was not on Hintikka's agenda. In $K \mathcal{C}B$ singular terms were recognized merely as variables and quantifiers were interpreted substitutionally; consequently, model set semantics did not involve any element, such as interpretation, denotation function, or variable assignment function, that could be seen as a formal counterpart of the language-world relation. After $K \mathcal{C}B$, Hintikka gradually became aware of this shortcoming. In Hintikka's new world-

¹² Cf. Chisholm (1976, 31) and Hintikka (1975, 17-19).

line framework, linguistic expressions refer successfully to objects by virtue of the intentional attitudes that we possess and, moreover, reference depends not merely on *what* we individuate but also *how* we individuate. Technically Hintikka deals with these intricacies of reference by means of quantification. Thus a thorough understanding of reference in epistemic contexts is crucial for understanding Hintikka's central formal tool, the two sets of quantifiers and their interpretation. Or is it the other way around? Perhaps a better way to put it is that once we have acquired Hintikka's system, a proper understanding of the preconditions of quantification is crucial for understanding reference in epistemic contexts. Nonetheless, reference and quantification are two sides of the same the coin, namely the phenomenon of individuation.

But how can that be? Reference, we are taught, is paradigmatically a feature of simple singular terms such as proper names, while quantified statements are statements that say of a predicate (such as *drinks*) that it is true of something (*Someone drinks*) or of everything (*Everybody drinks*) in a relevant domain (for instance, in a bar). Some quantified statements, definite descriptions, single out an object by virtue of their meaning but as Russell (1919, XVI) put it, strictly taken they do not refer.

An interesting philosophical question is what commitments and what limitations arise when Hintikka allows these notions to collapse into one another. As we shall see in Chapter 3, the result of this assimilation is neo-Kantianism: epistemic reality (or realities, possibilities) and *the* reality itself conflate in Hintikka's second framework, that is, we cannot distinguish the world as it is from our conceptual reconstruction of it, and hence there is no sharp distinction between epistemology and metaphysics. This defines a conceptual profile that is perhaps not unique to Hintikka's approach in contemporary philosophy of language, epistemology, and metaphysics, but it is marginal.

Rigidity

The characterization of rigidity given by Kripke (1980, 48) in *Naming and Necessity* states that an expression is a *rigid designator* if it designates the same object in every possible world in which the object exists, and a *non-rigid* or *accidental* designator if it does not designate the same object in every possible world.

One of the questions concerning rigidity in the context of this study is whether the content of that notion depends on the nature of the specific modality under discussion. In the case of alethic modalities, it seems reasonable to assume that the notion has ontological implications. Apparently, any expression may, in principle, be a rigid designator, even an expression designating a merely possible individual. But according to Kripke (1980, 156–158), rigid designators are meant to pick up objects from the domain of the actual world. If we distance ourselves from Kripke's *actualism* and tolerate merely possible objects that may be found not in the domain of the actual world but only in the domains of other possible worlds, we

face the following problem: if the object picked up by a rigid designator does not exist in the actual world then the *sameness* of the object in the other possible worlds becomes questionable. What would it mean that the same merely possible object, say, the brother I never had, is picked up in all possible worlds? Hence even a *possibilist* who allows merely possible individuals may want to restrict the designata of rigid terms to actual objects in order to make sense of the notion of sameness.

If the target modality is epistemic, what happens to these assumed ontological implications? In other words, if we posit rigid designators in a Hintikka-style framework with *epistemically* possible worlds, what does it mean that a designator picks out the same object in all possible worlds? Should we add a qualifier to the definition of rigidity and say that an *epistemically* rigid designator picks out the same *epistemic* object (or perhaps epistemically determined object) in all epistemically possible worlds? If we do add that qualifier, what does it exactly mean? Of course, this is just one symptom of our general bewilderment concerning the notion of epistemic possibility.

From a historical perspective, a book-length study could be written solely on Hintikka's extraordinary struggles with the notion of rigidity. Hintikka was among the first to introduce a notion of that kind and among the first to consider it fundamental in modal semantics.¹³ Then Hintikka went back and forth between rejecting and endorsing it. For instance, in 1995 he wrote a long critique (together with Sandu) of rigidity and "the new theorists of reference" attacking the notion from many viewpoints, but a few months later he had a new paper out in which he was quite happy to posit the class of rigid designators in his formal semantics under the name "proper constants", appropriately crediting Kripke for introducing such terms under a different name (Hintikka 1996, 122). His only complaint this time was that Kripke's rigid designators are "operational" in all worlds while his own proper constants "may fail to be operational in some world or other". It is not clear what Hintikka meant by this; it may have been a misunderstanding. A standard interpretation of a Kripkean rigid designator has it that it picks out the same individual in all worlds in which the individual exists, and if it does not exist - that is, if the individual in question is not to be found in the domain of some particular world – then the designator is not "operational" in the sense that it does not pick out anything from the domain of that particular world, as any textbook will tell you.

Some versions of two-dimensional semantics allow worlds in which Jupiter exists but the rigid designator *Jupiter* does not pick out Jupiter due to the differing meta-semantic facts prevailing in those worlds (Stalnaker 1981). This is meant to reflect the independence of meta-semantic facts from semantic facts. It is a semantic fact that *Jupiter* rigidly designates what it designates in all possible worlds, while it is a meta-semantic fact that *Jupiter* designates the largest gas giant of the

¹³ Hintikka's first remarks on rigidity may be found in Hintikka (1957a, 60).

Solar system to begin with. If things had been different, we might have used *Jupiter* to rigidly designate something else than Jupiter in all possible worlds. But now, things being as they are, we use *Jupiter* to rigidly designate Jupiter in all possible worlds. But it is unlikely that Hintikka was alluding to such aspects of rigidity. Hintikka (1996, 121–123) wrote that world-lines may break down, and by this he meant just that in some worlds there may be a gap in a world-line: that is, the function does not yield to a value in such a world if there is nothing suitable to pick out.¹⁴

Another curious claim made by Hintikka concerning rigidity is related to his aforementioned views concerning intuitions. Hintikka (1999) claimed that Kripke's idea of rigid designation was a whim, a hasty impression based on a quick intuition invoked by certain peculiar thought experiments such as the famous *Gödel* example, and not a rigorously developed position in the contexts of model theory and modal logic. I turn to these questions in Chapter 3 in which the main exposition on the topic of rigidity appears. However, due to Hintikka's complicated relationship with rigidity, the discussions relating to it are scattered throughout the chapters of this study.

All in all, Hintikka's writings on rigidity have an ambivalent quality. There is a tension between Hintikka's strict rejection of Kripke's formulation of rigidity and his acceptance of something very similar under a different name. This gives the impression that Hintikka was merely splitting hairs in his criticisms. In order to provide a proper alternative to Kripke's view, Hintikka should have resigned from Kripke's *realism* concerning the relational models in modal semantics. I argue that Hintikka should have stated clearly that he was an *instrumentalist* concerning the relational models.

Hintikka repeatedly wrote about the sameness of denoted objects being, for instance, a prerequisite for existential generalization. This creates the impression that, after all, Hintikka accepted the spirit if not the letter of the Kripkean view that, first, it makes sense to speak about objects as being the same from world to world and, second, that we possess the means to track those objects from one world to another. If we accept these points, then there is no room for a sweeping criticism of rigidity. I argue in Chapter 3 that it is not a criticism of rigidity to claim that, instead of there being rigid designators, there are physical world-lines carrying out the very same task, namely tracking objects themselves from world to world. It is mere cosmetics; it is replacing one terminology with another.

It is crucial to note that Hintikka's criticism of rigidity took place, as he put it, in the general context of the "realistic pragmatics of modal logic" (Hintikka 1975, 28). The original skeptical claim of Hintikka was *not* that proper names are not rigid, but a significantly stronger claim, namely that no singular term is rigid at the

¹⁴ This paper is a paradigmatic example of the recycling of ideas typical to Hintikka. The view put forth in this paper incorporates models sets, world-lines, and rigid designators into an eclectic mix.

outset and hence rigidity has no place, or is dispensable, in the semantics of applied modal languages.¹⁵

Propositions

What are our beliefs about? David Lewis (1979, 513) once listed objects of his beliefs; the list included his cat Bruce, winter, stormy weather, shoveling snow, fatigue, and that someday mankind will inhabit at least five planets. Lewis found this miscellaneous collection (apparently consisting of an animal, a season, a phenomenon, an activity, a mental state, and a state of affairs) too diverse and decided that it is better to stick with the traditional picture which assumes that objects of belief are uniform in ontological category. Hence, according to Lewis, the question of the objects of our beliefs should be answered not by listing concrete things that are in the world but by bringing up the stock notion of *proposition:* our beliefs are first and foremost about propositions and only derivatively about laptops and food and other interesting things laying around us. This view is as old as epistemology itself. So is the follow-up question: what are propositions?

The received general theoretical framework has it that propositions are things that sentences express, speakers assert, and correct translations preserve. We seem to commit ourselves to the existence of propositions by saying things such as *I'm offended by what you just said* and *All that would have been even more offensive if I'd said it in Russian*. Since propositions seem to be around anyway, why not consider them as objects of our beliefs and other epistemic attitudes? After all, the most interesting feature of propositions is that they are about or consist of the very same stuff that our beliefs are about, namely objects and their properties.

Perhaps one feature that makes propositions attractive is that most of the time one does not have to assume much about their exact nature. Merely by assuming that propositions are entities of some sort suitable for describing the epistemic attitudes of persons (as numbers are abstract entities suitable for describing many things) and capable of attributing properties truly to objects, one may lay down an impressive amount of philosophical theory to deal with many issues. For instance, one may deal with the following problems relating to objects of belief. First, it seems that we frequently entertain beliefs and other attitudes towards things that are unspecific or do not exist. The *list-view* that Lewis abandoned would require us to find and list things in the world that are unspecific and non-existent to serve as objects of such attitudes. Listing such things raises difficult ontological questions. By assigning objects of attitudes uniformly from the category of propositions, we

¹⁵ See e.g. Hintikka (1975, 28-29; also fn. 8).

¹⁶ As Prior (1971, 4) put it, the phrase *object of belief* is ambiguous. It may refer to what our belief is about or to what we believe. The last belief in Lewis's list was about mankind but the believed thing was the proposition *that someday mankind will inhabit at least five planets*.

¹⁷ The *lekton*-theory of the Stoics was one of the first systematic versions of this view, see Perälä (2014, 359).

avoid positing unspecific or indeterminate objects and avoid positing non-existent objects (whatever that means). If I want a yacht or suspect that the Merseyside stockbroker Marmaduke Bloggs is a character invented by the journalists, then, according to the propositionalist picture, these attitudes are not towards unspecific or non-existent objects but towards propositions involving general properties and relations.¹⁸ If I want a yacht, then my attitude is towards a general proposition which specifies the target of my desire as something having the property of being a yacht. More specifically, we may assume that there is a proposition that the intersection of the set of things having the property of being possessed by me and the set of things having the property of being a yacht is non-empty, and that I want that proposition to be true. Any object with both these properties would make the proposition true, and my desire would thereby be fulfilled. No unspecific yacht object (whatever that would be) is needed in the analysis. Similarly, there is no need for a non-existent Mr. Bloggs to serve as an object of my suspicion. There may be a general proposition that there is no object having the property of being a Merseyside stockbroker called Marmaduke Bloggs, and I might suspect that that proposition is true even though some journalists claim otherwise.

Second, beliefs stand in logical relation to one another – apparently because of the logical relations that prevail among the objects of beliefs. But a season and an activity (or any two contingent natural events for that matter) do not stand in a logical relation to one another in the way that beliefs about a season and an activity might stand. A simple explanation for this is that objects of beliefs are propositions that may stand in logical relations to one another, such as the following propositions:

If it is winter, then I have to shovel snow

It is winter;

Therefore: I have to shovel snow

This propositional approach brings in a strong linguistic emphasis in the study concerning the nature of belief: If the objects of belief are propositions then a belief is about something in the world by virtue of the structure of the proposition and certain items in the world. For example, the sentence *Bernard is hungry* expresses a proposition with a structure of two parts, one standing in a relation to Bernard and one to the property of being hungry. These parts are contributed to the proposition by linguistic expressions, that is, the name *Bernard* contributes a thing that determines Bernard, and the predicate *is hungry* contributes a thing that determines the property of being hungry. If we want to know more about the exact nature of the propositions and how they succeed in bringing the latest installments of consumer electronics and other worldly attractions to our attention,

-

¹⁸ These famous examples are originally from Russell (1905) and Grice (1969).

we must ask: what are these *things* that linguistic expressions contribute to propositions? Potential answers have a bearing on the broadest philosophical questions out there: what are the interconnections between language, world, and human thought? This study does not answer these questions. The primary topic here is not the metaphysics of propositions. Rather, it is to understand some of commitments that Hintikka's views have regarding the above issues.

Hintikka's first framework was *Ersatzist* and metalinguistic in spirit. In Hintikka's analysis possible worlds were sets of sentences reminiscent of Carnapian state-descriptions, and hence the objects of epistemic attitudes were not propositions but sentences. The resulting package was elegant. Linguistic *Ersatzism* pairs well with a modal metalinguistic theory of propositional attitudes by providing natural objects for epistemic attitudes, namely sentences. Hintikka was in a position to enjoy all the aforementioned benefits of the propositionalist theory without being committed to propositions as entities, which, of course, was a metaphysical relief. Despite this economy and elegance, Hintikka abandoned his metalinguistic approach. The reasons for this change are studied in Chapters 2 and 3.

Ontology/Epistemology Interface & Internalism/Externalism Divide

Trivially, there is a connection between language and world, and that connection somehow enables us to use sentences to say things about the world. There is also a connection between language and thought, and that enables us to use sentences to say what we think. Traditionally this setting has been pictured by the means of the so-called semantic triangle, which dates back to antiquity:

WORD



Figure 1 Semantic triangle

The two sides of the triangle may be taken to illustrate the two contrasting approaches to how language works and what propositions are. One approach stresses the right-hand side of the triangle connecting Word and Thought: the

things that linguistic expressions contribute to propositions are thoughts, and language thereby gets its representational capacities from Thought. The views that fall under this approach are customarily called "internalist" since the starting point of these views is the internal contents of our minds, which we find by introspection and reflection. The central question of intentionality within this approach is how to reach out from our conceptual sphere to the external world.

The contrasting approach emphasizes the left-hand side of the triangle connecting Word and Object. According to this approach language gets its content directly from the world: the things that expressions contribute to propositions are objects and properties in the world. The views that fall under this approach are customarily called "externalist", for their starting point is external to the mind: there are speakers of a language with complex physical constitutions in their environment, with certain behavioral capacities and causal connections to the world. But how do speakers' intentional internal states arise from this configuration? The strengths and weaknesses of internalist and externalist approaches may be illustrated in connection with the traditional demarcation concerning propositional attitudes, namely the distinction between the *de dicto* and the *de re* attitudes.

Bach (2010, 44) wrote that someone unknown to him "came up with the bad idea of extending the de re/de dicto distinction from modalities to attitudes", and as if that weren't bad enough, the same person supposedly "made matters worse by extending it to the attitude reports." I do not know who performed these deeds, but apparently they were committed before the debates between William Heytesbury and Jean Buridan. They raised questions such as whether Some star I know to be above our hemisphere follows without qualification from I know some star to be above our hemisphere. 19 The problem in the case of attitudes (not to be confused with the case of their reports or ascriptions) is that there is a difference between knowing that a language-related item (dictum) is true and knowing of some specific thing (res) that it has a certain property. This seems to commit us to two kinds of knowledge, and the propositionalist view is applicable only to the first kind, since the de re kind is supposed to be about an object, for instance about a particular star as in the above case, and not about a dictum. On the other hand, the traditional problem with Buridan's inference²⁰ has to do with reports or ascriptions (not to be confused with the problems about attitudes themselves): the latter report is ambiguous, as we do not know which attitude it reports. It might report a weak de dicto attitude that I know that there are stars above our hemisphere, or it might report a stronger de re attitude that I know of some particular star, say the Sun, that it is above our hemisphere. A further problem about reports was pointed out,

¹⁹ See Heytesbury's De scire et dubitare in Kretzmann & Stump (1988, 444), and Buridan's Summulae de Practica Sophismatum in Klima (2001, 901–902). Another early source is the treatise by "Pseudo-Scotus" to which Chisholm (1963, 774, 787–788) credits the distinction, but e.g. Knuuttila (2008) estimates that this treatise is later than the works of Buridan.

²⁰ The example is from Buridan's Summulae de Practica Sophismatum in Klima (2001, 901–902).

among others, by Kripke (1977, 103–104): the first report is also ambiguous. The report itself is a *de re* report but it does not automatically report a *de re* belief. Kaplan (1989, 555) has called these cases "pseudo *de re*".

As already mentioned, de dicto attitudes are mediated by some language related items while de re attitudes are in some special sense directly about concrete objects (or at least this is what the traditional understanding of this distinction suggests). The externalist Word and Object approach seems to be well placed to explain what this special directness of de re attitudes is, but in fact both approaches fall short in the case of de re attitudes. One way to look at the issue is that de re attitudes seem to connect objects and thoughts via the shortcut of the third side of triangle bypassing language altogether. However, this viewpoint is not available to the propositionalist approaches at hand, as the internalist and the externalist views here are both variants of the same theme, namely propositionalist semantics that aim to explain communication, competence, truth conditions, and basically all semantic phenomena by assigning propositions to sentences - and explaining epistemic attitudes as attitudes towards propositions. Since they are committed to propositions as abstract mediators of epistemic attitudes, they are forced to depict de re attitudes as connecting objects and thoughts by carrying objects from the left corner of the triangle over the top via language all the way down to the right corner to thoughts. The externalist approach focuses on the traffic on the lefthand side of the triangle and provides a picture of how names and predicates contribute objects and properties themselves to de re propositions. However, the picture does not carry over to the other side since this approach lacks immediate means to describe the significance of de re propositions to a thinking subject.

The internalist Word and Thought approach, on the other hand, looks well placed to explain the relation and the significance that language bears to a thinking subject. But according to it, names and predicates always contribute thoughts about objects (and properties) to propositions. Therefore the *de re* propositions involved in the *de re* attitudes cannot involve objects and properties themselves but merely thoughts about them. Since a thought about an object is still just a thought, the approach confronts the same problem of being stuck on its own side of the triangle and fails to say interesting things about the supposed special object-bound nature of *de re* attitudes.²¹

A predictable feature of the recent development of these views is that the externalists²² have resorted to thoughts under some technical term such as "guise", "a way of grasping", or "a way of believing a proposition" in their analyses, while many internalists²³ have resorted to objects by invoking linguistic expressions that conveniently contain objects themselves. The externalists of this kind hold that a

²¹ The internalists are more subtle on this. For instance, Frege did not think that the contents of thoughts are themselves thoughts in this sense: they must be something more objective, something that the speakers can share.

²² See e.g. Salmon (1986, Ch. 8); Soames (2002, Ch. 8 & 3); Thau (2002, Ch. 4); Braun (1998, 2002).

²³ Bach (1987, Ch. 7), Jackson (1998b), Chalmers (2011).

given sentence always expresses a proposition containing objects and properties themselves, but when the sentence is used it comes with a built-in thought. The internalists of this sort hold that a given sentence always expresses a proposition containing things that we think, but when the sentence is used it may come with built-in objects. Hence the two approaches, an ontologically driven theory and an epistemologically driven one, have become very similar despite their radically different starting points.

Does Hintikka's approach provide a distinctive take on these matters? We shall see that, after abandoning model sets, Hintikka seems to take a propositionalist stand: all attitudes de dicto and de re are similar in the sense that they are directed towards propositions that in turn are analyzed as sets of worlds. We might also suggest that Hintikka's framework provides a way to bypass the semantics of natural language in the sense that we may, as a first step, represent the information that the agent possesses in a given scenario without linguistic concerns, and after that, as a second step in the analysis, consider the role that the linguistic expressions involved must play relative to the possessed information and epistemic attitudes. In principle this analysis could be reversed. We could first lay out some semantic views and state, for instance, that all proper names are rigid and then produce some problematic statements, and then, as a second step, draw conclusions, say, about the consistency of certain epistemic attitudes concerning the statements. This turnaround, however, is not generally available to Hintikka. The core idea of Hintikka's approach is to analyze ascriptions of propositional attitudes by quantifying over possibilities consistent with what the epistemic subject knows or believes. If the subject holds contradictory attitudes then there would be no relevant possibilia available: no contradiction is true in any possibilia.²⁴ Sandu (2006, 543) wrote, and Hintikka (2006, 555) approvingly quoted the declaration, that "for Hintikka, matters of reference are always intermingled [...] with cognitive matters". True, but strictly taken the "matters of reference" are not merely intermingled but always subordinate to and governed by the matters of cognition in a Hintikkian analysis, precisely in the above sense. The matters of reference may be assessed only after coherent epistemic attitudes are established. This is one of the features that make Hintikka's approach internalistic and undermine its capability to break out of the epistemic sphere. I spell out this point more specifically in relation to *de re* attitudes in the next section.

Hintikka claims in many of his writings that his frameworks are to be understood as neutral tools that are suitable for different purposes and largely

²⁴ Many authors, especially linguists, follow Kratzer here and consider Hintikka's approach a failure in this respect, see e.g. Hegarty (2016, 9–10) and the papers collected in Kratzer (2012). According to this view, inconsistent attitudes are naturally occurring phenomena and a theory of propositional attitudes must take them into account. In the case of possible world analysis this means that the relevant *possibilia* must be determined without appealing to consistency. Hintikka might have replied by appealing to the impossible possible worlds at play or by taking an Aristotelian stance similar to Barcan Marcus' (1981, 1983).

independent of popular philosophical positions.²⁵ This is tendentious. Hintikka's epistemic internalism, for instance, is evident in the light of many his remarks. 26 His semantic internalism, in turn, is not so obvious. In general, Hintikka left many questions open in his philosophy of language; he was vague and stressed (especially in his early writings) that he was a logician making logical distinctions when, in fact, he was presenting and especially presupposing substantial claims about meanings, metaphysics, and human cognition (see e.g. Hintikka 1969, 169-170; Hintikka 1969, 108). For instance, Hintikka's most important discussions of meaning contained casual remarks on how his semantic framework was compatible with and superior to Frege's, Quine's, Kaplan's, Montague's, and Kripke's famous ideas but, as expected, the details were omitted.²⁷ Hintikka (1969, 87-88; 180) also used to emphasize the exclusive role of reference in his semantics while, perhaps paradoxically, endorsing some reading of Fregeanism (a reading whose specifics have remained veiled). As we shall see in Chapter 3 this tension culminated when Hintikka stressed the importance of both "rigid" reference and descriptive modes of individuation in a manner that comes across as contradictory: The referentialist anti-Fregean doctrines on the exclusive role of reference in semantics do not mix with the indispensability of "modes of individuation", entities clearly reminiscent of Fregean senses.

De Re Attitudes

How, then, should we understand the details of *de re* attitudes, and especially the details of *de re* attitude ascriptions against the theoretical stances characterized above? The most popular strategy is reductive: the ascriptions of *de re* attitudes should be understood in terms of *de dicto* attitude ascriptions, that is, all object-dependent attitudes should be understood as proposition-dependent attitudes, at least when ascribed to agents.

Quine (1956) and Kaplan (1968) provided detailed frameworks for reducing *de re* belief ascriptions to *de dicto* belief ascriptions. The main problem in such a reduction is that *de re* belief ascriptions seem too sparse from the propositional point of view. A *de re* construction merely sets some external constraints on the believed proposition which (in the light of Quine's examples) can be rich in conceptual and representational content. A harsh *de re* ascription in propositional terms ascribes to a subject a belief such as

x is a spy

-

²⁵ See e.g. Hintikka (1962, 20); Hintikka (1969, 169–170).

²⁶ Hintikka's remarks on "adequate" justification, "strong" epistemic notions and their "conclusive grounds", and his defense of the so-called *KK-principle*, i.e. the principle that if one knows that *p* then one knows that one knows that *p*, are clearly more compatible with epistemic internalism than any externalist position (Hintikka 1962, 20–21); Hintikka (1970, 145–146).

²⁷ See e.g. Hintikka (1969, 180), Hintikka (1962, 150–155), Hintikka (1975, 73–74); Hintikka & Sandu (1995, 250–252), Hintikka (1998, 229).

with x standing in for some specific object. The problem with this is that no-one seems to believe such a minimal thing. As Hintikka (1957a, 50) put it, "one may ask, who was this x anyway?". Nobody believes that a certain object is a spy independently of how that object is referred to, that is, without associating some attribute with that object. What can be believed is something along the lines of The man on the beach is a spy or The person known as 'Bernard' is a spy. Quine argued that this associated content must be recovered for a successful reduction of de re belief ascriptions to de dicto belief ascriptions. Both Quine and Kaplan posited additional conceptual or representational elements in their analyses of de re beliefs in order to re-establish them in propositional terms. Quine (1956, IV), at the first stage of his analysis, posited arbitrary attributes to enrich beliefs such as 'x is a spy', which he called relational beliefs and then, at the second stage, tried to eliminate them with metalinguistic predicates that apply to sentences and predicates. Then he modeled the de dicto belief (or the nominal belief as he called them) as a belief about a sentence that it is true while the de re belief (or the relational belief) came out as a special instance of this metalinguistic attitude: to believe something de re, about an object, is to believe of a predicate that it is satisfied by an object. Kaplan (1968, IX), in turn, introduced the notion of vividness. A vivid term enters the analysis with a denotation-determining sense that adds conceptual content to de re ascriptions. These enhancements are in line with a Fregean view. The common background sentiment shared by these theorists is that an expression inside a belief context always contributes more than merely its referent to the overall meaning of a successful attitude ascription. In addition to providing the referent, it must provide some attribute of the referent. A Fregean offspring of this view is the aforementioned program of epistemic two-dimensional semantics, which distinguishes two dimensions of meaning in the linguistic analysis of belief.²⁸

The need to enrich the analysis of *de re* attitudes with additional conceptual elements arises especially when we have a Fregean conceptual and representational view on the nature of propositions and beliefs. According to the rival neo-Russellian view, *de re* attitude ascriptions are not sparse with respect to propositions. Since Russellian singular propositions are composed of objects and properties themselves, there is nothing "representational" or "conceptual" missing from *de re* constructions from the propositional viewpoint. The propositional consensus can be maintained by holding that to believe *de re* is to believe *de dicto* a Russellian singular proposition. In the contemporary philosophy of language, the Russellian understanding of propositions and an externalist (anti-representational and anti-conceptual) emphasis on the nature of beliefs are implicit in the theories of direct reference. Kripke's (1980, I & II) examples of direct reference determination with no conceptual or representational mediators involved in

²⁸ See e.g. Jackson (1998a, 1998b); Chalmers & Jackson (2001); Chalmers (2002a, 2002b, 2004, 2006, 2009, 2011).

Naming and Necessity can be read as defending sparse de re attributions as reasonable, since in most cases believers do not associate rich conceptual or representational contents with the things that their beliefs are about.

Sometimes, on the other hand, the believers and the knowers obviously do associate contents with objects of their attitudes. Especially in cases depicted by the famous belief and knowledge puzzles, the associated content seems relevant to the attitudes themselves and to the semantics of attitude ascriptions. Consider an extreme case: sometimes it might make good sense to say that I have x as an object in my mind in the sense that I can somehow refer to it or bring it to attention by, let us say, using a completely incorrect description. But to say that I have knowledge about x seems to be a different thing. For then Hintikka's question arises: what is this x anyway? I cannnot know about something I do not know anything about.

As mentioned in the previous section, Hintikka has throughout the years emphasized in his writings that his theories are strictly anti-Fregean, making no use of "[...] Fregean senses, David Kaplan's characters, Montague's meaning functions or other such paraphernalia". 29 I argue that both Hintikka's frameworks involve a Fregean internalist and representationalist conception of epistemic attitudes. According to Hintikka's model set semantics truth was a matter of mere coherence, all modality was conceptual, and the connection between symbols of language and objects in the world manifested only in a consensus between the language-users, and consequently our cognitive access to reality and eventually, knowledge, was a matter of consensus. Hintikka's second world-line framework, in turn, was a neo-Kantian program developed against a background view that even in order to perceive an object (not to mention other attitudes) we must apply a rich arsenal of conceptual, representational, and linguistic resources such as quantification, cross-reference, and cross-identification. Moreover, in a Kantian spirit, these resources do not provide us epistemic access to the world as it is but merely access to our own reconstruction of it. The conceptual arsenal entering the analysis isolates epistemic agents from the reality itself into an epistemic sphere and hence ontological viewpoints collapse into epistemological viewpoints. When we ascribe an attitude towards this or that object in Hintikka's formalism, the conceptual arsenal does not guarantee that the ascribed attitude is about a particular object. It merely guarantees that the ascribed attitude is about some object or other that the agent takes to be in the extension of accompanying predicates, be it one-placed or two-placed, such as identity. Hintikka's objects are neo-Kantian reconstructions; from the viewpoint of the agent they may be indistinguishable from the objects themselves, and some form of epistemic referential stability may perhaps be defined for terms referring to these epistemic objects but, at the end of the day, Hintikka's objects are not objects themselves.

²⁹ Hintikka (1998, 229). For other similar statements see e.g. Hintikka (1969, 87-88, 180).

The objects themselves come out as transcendental and therefore Hintikka's formal epistemology, as it stands, lacks the resources to accommodate *de re* attitudes as they are depicted by the neo-Russellian views.

One obvious obstacle in ascribing *de re* attitudes is that Hintikka's formalism lacks a device suitable for ascribing epistemic attitudes towards singular propositions. The basic *de re* attitudes were meant to be captured by the formulae such as

$$\exists x K_s(b = x)$$

with the intended reading "S knows who/what b is". But these formulae were not suitable for carrying out that task because the early model set semantics was too economic to support the meaning that Hintikka intuitively assigned to them, and in the case of the world-line framework Hintikka's neo-Kantian presuppositions prevented them from guaranteeing that the same metaphysical object is picked out in all epistemically possible worlds, let alone in metaphysically possible worlds. The new semantic story that Hintikka (1969, 171-172; 1975, 47-49; 2007) provided for formulae such as $\exists x K_s(b = x)$ was that their truth depends on there being a physical world-line reflecting the fact that the subject S "operates" in a physical mode of individuation (which tracks objective physical properties and regularities) relative to an object, and that this object is known by the subject S as 'b' in every relevant possible world. True, the interpretation of the formula guarantees that there is some b in all relevant possible worlds but it does not guarantee that b is the same object in these worlds. It merely guarantees that the physical properties and regularities by virtue of which b was individuated in the actual world are to be found as attached to some object or other in all relevant possible worlds, and that this object is known as 'b'. This attitude is not a de re attitude or the knowing who attitude described in K&B. The original demarcating feature of such an attitude was that it is directed towards a set of worlds in which the individual that the attitude is about is the same, and all worlds occupied by some other individuals are eliminated as incompatible with the original de re attitude.

It is important to note that if we take the strict instrumentalist neo-Kantian position as characterized in Chapter 3, then it is impossible to ascribe proper *de re* attitudes to the subject since there would be no suitable *res* available that the attitudes could be directed to. A proper *de re* attitude, as defined for instance by Kaplan (1989, 484) and Salmon (1986, 3–7), is one that is dependent on and determined by the thing that the attitude is about. The object in the world determines the singular proposition, which in turn determines the attitude as *de re*. But this is not possible in Hintikka's frameworks. The direction of determination is wrong. Frege's famous view implied that there is no road back from the referent to *Simn*. Similarly there is no road back for Hintikka from the object itself to the proposition and to the attitude. We cannot pick out objects and determine some

attitudes that are directed to them as $de\ re$. In order to pick out objects, we must take into account the underlying conceptual arsenal, which determines the attitudes that determine the propositions that pick out the objects. Such ascribed attitudes can never be semantically $de\ re$ in the above sense – unless we allow Hintikka's $de\ re$ quantification collapse into rigidity and let $\exists x K_s(b=x)$ just mean that 'b' is rigid. Chapter 2 argues that this is not an option in the model set framework, and Chapter 3 studies whether this is a viable option in world-line semantics.

This vagueness of *de re* attitude ascriptions casts serious doubts on the relevance of Hintikka's epistemology: for instance, the analysis of *knowing who* ascriptions becomes questionable. *De re* attitudes are simply too important, no matter what one thinks about direct reference or singular propositions. An almost poetic expression of the widely shared sentiment can be found from Burge, a well-known critic of the direct reference orthodoxy. He wrote in 2009 that

[...] understanding *de re* phenomena is a project not only in the theory of reference, let alone belief-attribution. It is a project that probes fundamental epistemic and representational capacities that underlie what it is to have a mind. (Burge 2009, 316)

It is surprising that Hintikka's undeniable problems in accommodating *de re* attitudes (and the specific reasons for them) have not been clearly stated in the literature before. The most straightforward recognition comes from Hintikka himself. In his aforementioned 1996 paper on *proper constants* (rigid designators) he wrote:

In earlier books and papers, Jaakko Hintikka has dealt with the problem of knowing who *de dicto*, we might say. What we are doing here is to give an analysis of *knowing* + *wh*-construction in its *de re* sense. (Hintikka 1996, 126)

The statement is clear: an account of *de re* attitudes was not possible before the introduction of rigid designators.

Dimensions in Modal Semantics

One way to survey some relevant variants of 2D semantics for the purposes of this study is to try to evaluate their level of internalism in the sense characterized above. Stalnaker's (2004; 2008) discussions provide a good starting point for such a task.

Stalnaker's (2004) 2D assigns to certain terms such as *water* two sorts of semantic values, or perhaps one might say: a semantic value on the other hand and a "meta-semantic value" on the other. The semantic value of *water* is an ordinary

intension (standardly associated with rigid designators in the formal semantics of modal languages) understood as a constant function from possible worlds to extensions determining the same substance, H₂O, in all possible worlds as the extension of the term. The meta-semantic value is a two-dimensional intension understood as a function from possible worlds to semantic values that are themselves functions, that is, ordinary intensions. Hence, intuitively, the twodimensional intension determines a set of worlds in which the term water has some semantic value or another, and then in each world an ordinary intension determines what extension the term has in that world and, consequently, in all possible worlds. This is meant to reflect the independence of meta-semantic facts from semantic facts. It is a semantic fact that water rigidly designates what it designates in all possible worlds, while it is a meta-semantic fact that water designates H₂O to begin with. If things had been radically different, as they appear to be in Twin Earth, in which rivers, lakes and oceans are filled with XYZ, we might have used water to rigidly designate something other than H₂O. But now as things are as they are, we use *water* to rigidly designate H₂O in all possible worlds. Stalnaker's main applications of this framework concern situations in which a speaker is aware of the semantic fact that a certain term rigidly designates whatever it designates in all possible worlds but remains ignorant of the meta-semantic fact of what exactly is the designated thing. Roughly, such a speaker cannot be informed of the meta-semantic fact by means of the statement containing expressions with ordinary intensions assigned as their semantic values. A statement containing expressions with two-dimensional intensions assigned as their "metasemantic values", however, may be appropriate for such a purpose.

The level of internalism in Stalnaker's interpretation is close to zero since meaning, intentionality, and cognitive access to reality are phenomena that take place in the relation to the world and the prevailing facts therein, that is, in the realm of semantic values understood as ordinary intensions. According to Stalnaker's externalist position, it is the nature of H₂O and the causal connections we bear to it that explain why the term *mater* has the semantic role in our statements that it has. In possible worlds where the chemical facts are different, the externally determined semantic facts are also different. If the nature of the stuff in rivers, lakes and oceans were different, the causal connections would also be different, and this explains why the semantic value of *mater* could be different from what it actually is.

Stalnaker's interpretation of 2D framework is not designed to address the epistemic and cognitive questions Hintikka sought to address. Hence I focus on 2D frameworks that are epistemically and internalistically interpreted, such as David Chalmers' epistemic 2D. Chalmers has been developing his 2D framework for years in order to capture certain epistemic and cognitive aspects of meanings, and it is expected to reach its full maturity in Chalmers' forthcoming book *The Multiplicity of Meaning*. In addition to a generous amount of critical commentary,

Chalmers' program has spawned and inspired a number of surveys on the history of 2D approaches in semantics and logic. For instance, Jason Stanley (2014) evaluates the place and the state of Chalmers' program by stating "Chalmers' view of content has a major advantage over that of neo-Fregeans such as Evans and Peacocke". According to Stanley, Chalmers' framework provides Fregean senses for our disposal in a semantic theory while avoiding the problem concerning their metaphysical status, i.e. avoiding "[...] the charge of obscurantism typically levelled against the notion of Fregean sense" by introducing senses as "epistemic intensions" that are technically functions from linguistically constructed possible worlds to extensions in a modal system governed by Carnapian state-description semantics. In such an Ersatz-type modal approach possible worlds are sets of sentences, and at least in Stanley's view such sets and functions from them to extensions are metaphysically unproblematic. I do not evaluate this metaphysical thesis here. But I argue that Hintikka's work anticipated such ideas. For instance, Chalmers' neo-Fregean view of content that Stanley praises was spelled out in Hintikka (1969). Still, Hintikka's work has been totally ignored in the literature on two-dimensionalism. Chapter 4 assesses Chalmers' framework and compares it to Hintikka's work.

Introduction

2 CREATION OF EPISTEMIC SPACE

Who invented possible worlds semantics? Perhaps the most plausible answer is that many people did, over a long period of time. Contemporary possible worlds semantics is a conglomeration of philosophical ideas (such as understanding modalities as quantifications over worlds), technical innovations (such as accessibility relations), and mathematical results (such as completeness proofs). Many scholars have contributed, some independently and some dependently of each other, some more philosophically and some more technically. Some have contributed outstandingly in every aspect, for instance Saul Kripke, whose work relates especially to the technical and philosophical aspects of the type of modal spaces this study is concerned with, namely spaces with worlds containing individual objects. The scientific community honours his contributions by calling this type of relational possible worlds semantics "Kripke semantics".

Some consider this practise historically misleading. It has been frequently pointed out that many others deserve credit for Kripke semantics in addition to Kripke, for instance Tarski, Jónsson, Montague, Kanger, Guillaume, McKinsey, Bayart, and Smiley.³⁰ Jaakko Hintikka's name has also been mentioned frequently in this connection. Some have gone further and claimed explicitly that Hintikka was the first to introduce relational possible worlds semantics, or Kripke semantics, some years before Kripke. This claim raises many questions, beginning from its exact meaning.

Taken at face value, the claim is false. When we look at Kripke's semantics we find, for instance, set of possible worlds, domain of individuals, assignments of semantic values, and models. In Hintikka's modal semantics we find none of these. Instead, we find model sets, model systems, rules, and conditions. So it seems that we have two different semantic frameworks. Hintikka might have been earlier, but his semantics was different from Kripke's.

The claim, however, is more subtle: despite their differences the two frameworks reflect the same philosophical ideas and provide essentially the same semantical analysis of quantified languages involving modalities. After all, both approaches provide definitions of semantic notions of *validity* and *satisfiability* of modal statements in terms of *alternative states of affairs* or some closely related concept. When we put it this way – without highlighting the controversial question of priority – it turns out that the claim is widely accepted, at least implicitly: When Hintikka's early work is discussed in the contemporary literature, his ideas are usually presented by means of Kripke semantics. This practise, when unqualified, suggests that there is an unproblematic translation from Hintikka's

41

³⁰ The list of people present at the Helsinki colloquium in modal and many-value logics in the summer of 1962 may be also taken up in this connection: Tarski, Kripke, Curry, Montague, Kanger, Anderson, von Wright, Prior, Barcan Marcus, Chang, Hintikka, Ajdukiewicz, Smiley, Lemmon, Geach, Rescher, etc. (see *Acta Philosophica Fennica* 16).

early model set procedure to Kripke semantic treatment. If so, it seems that we have one semantic analysis spelled out in two different ways in two different vocabularies, and the choice between the two vocabularies is merely a matter of convenience.³¹ In 2006 Hintikka regretted that he had chosen the vocabulary of model sets:

Right from the beginning I was interested primarily in the semantics and pragmatics of modal and intensional concepts, as witnessed by my 1957 paper "Modality as Referential Multiplicity". For this reason, it was probably a mistake for me to use as the framework of presentation in my early studies the quasi-syntactical technique of model sets and model systems. It seems to have created in some people the impression that I was not trying to get at the actual model theory of modal notions. (Hintikka 2006, 20)

The fact is that the details of Hintikka's model set framework have remained poorly understood. Sections 2.1 – 2.8 of this Chapter assess the historical question of how Hintikka's epistemic modal space grew out of deontic and alethic frameworks, and survey Hintikka's early modal semantics in relation to Kripke's semantics for quantified modal languages. After assessing the relevant material I conclude that Hintikka's early model set semantics and Kripke semantics are, in fact, very different, and Hintikka's semantics cannot be understood as a variation of Kripke semantics. The main reason for this is that Hintikka's model set framework was built on syntactic foundations and it simply lacks the relevant semantic features. The last section, 2.10, evaluates prospects to develop Hintikka's framework by enriching model sets with sufficient semantic detail, as suggested in Hintikka's informal remarks.

Before proceeding, I wish to stress that my interest here is *not* the general historical question regarding the inventor or inventors of possible worlds semantics. My question here is the exact nature of model set semantics and its relation to Kripke semantics in the light of the *published* writings of the authors. Other possible points of view found in historical studies on the topic include, for instance, Copeland's (2002): he chooses to speculate on the content of Hintikka's *Boston Manuscript*, which has been lost for decades, and concludes that it contained semantics and a completeness proof similar to Kripke (1959b). That said, I have also done some detective work and I have learned that this manuscript was brought from Boston to Helsinki. I know people who have read parts of it and in my spare time I too have tried to locate it, without success, in libraries and in obscure depots and impromptu archives at the Helsinki department. But this detective work or the content of that manuscript or some piece of folklore that

³¹ This conception is put forth, for example, in *Wikipedia*: The entry "Kripke semantics" (June 2021) tells us "Jaakko Hintikka gave a semantics in his papers introducing epistemic logic that is a simple variation of Kripke's semantics [...]". According to the revision history of the entry this claim has been online since 2004.

one might learn in the corridors of the Helsinki department do not have any bearing on the issues studied here. The chapter is partly based on the previously published papers Tanninen (2016) and Sandu & Tanninen (2017).

2.1 KRIPKE SEMANTICS BEFORE KRIPKE?

Hintikka has claimed that his early semantic framework utilizing model sets comprised the central features of relational possible worlds semantics, or Kripke semantics, for quantified modal languages. Moreover,

the semantics envisaged in Hintikka's 1957[a] paper is [...] "Kripke semantics" five years before Kripke. (Hintikka 1982, 93)

Hintikka continued in the footnote that

This semantics was originally discovered independently by Kanger, Hintikka, Guillaume, and Montague and possibly still other logicians. Later, it was independently discovered also by Kripke and E. W. Beth. (Hintikka 1982, 103, fn. 2)

Kripke's own first-hand remarks in his early papers seem to confirm, at least partly, Hintikka's claims: in the 1959 abstract of a paper that eventually was published in 1963 in two parts, "Semantical Analysis of Modal Logic I & II", Kripke stated that he would provide completeness theorems for various systems of quantified modal logic, first, by defining model-theoretic semantics and, second, by proving the completeness of each system in relation to the defined semantics with the aid of *semantic tableux* introduced in Beth (1955). According to Kripke

For systems based on S4, S5, and M, similar work has been done independently and at an earlier date by K. J. J. Hintikka. (Kripke 1959a, 324)

To this Kripke added that

The resulting semantical notions shed new light on questions such as the morning star paradox, and provide a semantical apparatus for sense and denotation, extension and intension, and related concepts. (Kripke 1959a, 324)

In the published full paper Kripke stated that

For systems based on S4 and M and [...] on S5, HINTIKKA has discovered a modeling similar to the present one [capitalization in the original text]. (Kripke 1963a, 69, fn. 2)

In "Semantical Considerations on Modal Logic" Kripke wrote that

The [semantical] theory given here has points of contact with many authors [...] The authors closest to the present theory appear to be Hintikka and Kanger. (Kripke 1963b, 83, fn. 1)

Three explicit claims may be extracted from these remarks: 1) Hintikka's semantics for quantified modal languages, introduced in 1957, was essentially the same that Kripke gave, first in 1959 and then, with some significant modifications, in 1963; 2) Prior to 1959, Hintikka had proved completeness theorems for various quantified systems based on C. I. Lewis' S4, S5, and M; 3) Hintikka's semantics is applicable to philosophical problems that Kripke sought to address, namely to problems relating to semantic notions of sense and denotation and intension and extension.

I argue, first, that the semantics introduced and utilized in Hintikka's writings spanning from 1957 to 1968 is not Kripke semantics. Second, Hintikka did not publish nor had at his disposal Kripke-style soundness or completeness proofs for quantified S4, S5, M, or any other system: when it comes to proofs similar to Kripke's, it would have been simply impossible for Hintikka to provide such proofs because Hintikka, as we shall see, did not define (and was not interested in defining) model-theoretic semantics for the aforementioned systems (or any other system before his post-*K&B* papers of the late 60s). After all, Kripke's completeness result is about the relationship between certain axioms and a particular model-theoretic semantics. Strictly taken Hintikka did not introduce axioms and he did not define model-theoretical semantics, and therefore the question of their exact relationship cannot even emerge in his approach. Third, Hintikka's semantic apparatus, model set semantics, cannot accommodate the kind of distinctions between sense and denotation or intension and extension that Kripke referred to.

I proceed with a short note that lists some historical studies that emphasize Hintikka's role in the creation of the type of semantics under consideration. Then, since (i) the task at hand is a historical comparison and (ii) the framework put forth especially in Kripke (1959b) is quite different from any of the contemporary presentations of Kripke semantics, I give an outline of Kripke's early expositions in section 2.3. Sections 2.4 – 2.7 comprise an extensive survey of Hintikka's early writings, and I continue with a comparison of Kripke's and Hintikka's theories, including comparisons of their treatments of the Barcan formula and its converse in section 2.8. Sections 2.9 and 2.10 assess the difficult question of *quantifying-in* and possibilities to introduce Kripke semantic elements to Hintikka's framework. I

close with final critical remarks concerning the notions of completeness and soundness in Hintikka's early modal logics.

2.2 HISTORICAL STUDIES ON MODAL SEMANTICS

In general, historical studies on the development of quantified modal logic recognize Hintikka's role as pivotal in the creation of relational possible worlds semantics for quantified languages, see e.g. Goldblatt (2006) and Lindström & Segerberg (2007). Some go further, and take an explicit stance regarding priority. For instance, Copeland (2002; 2006) evaluates that Hintikka and Kripke together were the first "in a glorious photo-finish" to provide completeness proofs for the most well-known quantified systems (Copeland 2002, 100). According to Smith

Hintikka was the first to *develop* [...] and the first to *publish* [...] a *metaphysical* semantics for modal logic in terms of the reflexive, transitive, etc., features of the alternativeness relation. (Smith 1998, 259)

Føllesdal (1994; 2016), in turn, has argued that it is a mistake to talk about "Kripke semantics" as Kanger and Hintikka introduced its main ideas, both before Kripke (independently of each other and independently of Kripke).

There is also a "national concern". Throughout his international career Hintikka had a close connection with the Helsinki department and hence there are studies on his work available (at least currently) only in Finnish. A number of Finnish philosophers have argued in print that Hintikka provided the central ideas of relational possible worlds semantics before Kripke, most notably Halonen (2009) and Rantala & Virtanen (1996). I do not discuss the details of these studies here. I merely mention them to note that a number of studies *explicitly* argue that Hintikka introduced Kripke semantics before Kripke. As already mentioned, a majority of contemporary expositions of Hintikka's early work offer *implicit* support for this claim by resorting to Kripke models instead of model sets when discussing Hintikka's ideas, see for instance papers collected in van Ditmarsch & Sandu (2018), Lihoreau & Rebuschi (2014), and the 2019 Hintikka memorial issue of *Logica Universalis* (13/2). Now I turn to the exact relationship between Kripke's and Hintikka's semantic theories.

2.3 KRIPKE MODELS

Kripke (1959b) considers a modal language that contains individual variables x, y, z, ..., n-adic predicate variables P^n , Q^n , R^n , ... (which are also said to be propositional variables when n=0), the truth-functional logical connectives with identity, quantifiers, and the modal operators of necessity and possibility, \square and

 \diamondsuit (the latter is defined as $\neg \Box \neg$). The language does not contain individual constants. Given a non-empty domain **D** of individuals, for each formula A in the object language, Kripke defines the notion of a complete assignment for A in **D**, which is a function that assigns

- to every free individual variable of A an individual in \mathbf{D}
- to every propositional variable which is a sub-formula of $\mathcal A$ either the truth-value $\mathbf T$ or $\mathbf F$
- to every n-place predicate variable P^n occurring in A a set of ordered n-tuples of members of \mathbf{D} .

A *model* of A in \mathbf{D} is defined as an ordered pair $\langle \mathbf{G}, \mathbf{K} \rangle$ of complete assignments for A in \mathbf{D} , where $\mathbf{G} \in \mathbf{K}$ and all the assignments of \mathbf{K} agree on the assignments of free individual variables of A. Intuitively, the assignment \mathbf{G} plays the role of the actual world and the set \mathbf{K} is to be thought of as the set of all possible worlds, the modal space. There is no accessibility relation on \mathbf{K} . Given a model $\langle \mathbf{G}, \mathbf{K} \rangle$ for A in \mathbf{D} , every sub-formula B of A receives the value \mathbf{T} or \mathbf{F} relative to an arbitrary assignment $\mathbf{H} \in \mathbf{K}$ in a recursive way:

- (i) If *B* is an atomic formula $P(x_1, ..., x_n)$, then it receives the value **T** if and only if the *n*-tuple $(a_1, ..., a_n)$ assigned by **H** to the free individual variables $x_1, ..., x_n$ belongs to the extension of *P* as given by **H**; otherwise it receives the value **F**.
- (ii) If B is $x_1 = x_2$, then it receives the value **T** if and only if the individual in **D** assigned to x_1 by **H** is the same individual as that assigned to x_2 . Otherwise it receives the value **F**.
- (iii) The clauses for the common logical connectives are standard.
- (iv) If B is $\forall x C x$, then it receives the value T if and only if Cx is assigned the value T for every assignment of an element of D to x; otherwise it receives the value F.
- (v) If B is □C, then it receives the value T if and only if every member of K assigns the value T to C; otherwise it receives the value F.

The formula A is valid in a model $\langle \mathbf{G}, \mathbf{K} \rangle$ of A in \mathbf{D} if and only if A is assigned the value \mathbf{T} by \mathbf{G} (intuitively: A is true in the actual world). A is valid in \mathbf{D} if and only if A is valid in every model of A in \mathbf{D} . A is satisfiable if and only if there is a non-empty domain \mathbf{D} and a model of A in \mathbf{D} such that A is valid in this model. Finally, A is universally valid if and only if A is valid in every non-empty domain \mathbf{D} . B is semantically entailed by A_1, A_2, \ldots, A_n if and only if $(A_1 \wedge A_2 \wedge \ldots A_n) \to B$ is universally valid. If n = 0 this amounts to B being universally valid.

³² In Kripke (1963b) it is acknowledged that this terminology was not ideal: it would have had been better to state *true in a model* instead of *valid in a model* in this particular case.

Kripke proves a number of theorems, including "THEOREM 1", which shows that B is semantically entailed by $A_1, A_2, ..., A_n$ if and only if a two-column tableau construction (introduced in Beth 1955) where $A_1, A_2, ..., A_n$ are on the left side and B is on the right side of the tableau closes. This and the other theorems establish the soundness and completeness of quantified S5 with identity relative to the semantics above, that is, they form a proof that eventually "equates the syntactical notion of provability and the semantic notion of validity" (Kripke 1963b, 84). The system assessed by Kripke consists of a Hilbert-style axiomatization of first-order logic (with identity) supplemented by the modal axiom schemes known as M or T and the Distribution Axiom

$$\Box A \to A$$
$$\Box (A \to B) \to (\Box A \to \Box B)$$

and the axiom scheme characteristic of S5

$$\neg \Box A \rightarrow \Box \neg \Box A$$

The rules of inference are *modus ponens* and necessitation. The details of Kripke's proof are not relevant here. I will briefly return to these issues in sections 2.5 and 2.8. For now, let me emphasize a few points in Kripke's semantics that are relevant to the comparison with Hintikka, as we shall witness later:

- Semantics proceeds by defining the notion of *model*.
- A model assigns semantic values from D to linguistic items. The key semantic notions of *satisfiability* and *validity* are defined in terms of assigned values.
- Possible worlds are, or are represented by, complete assignments of semantic values.
- Possible worlds all share a common domain of individuals **D**.
- Variables are *rigid* in the following sense: a free variable x is assigned an element of the commonly shared domain **D** and the members of **K** are all assumed to agree on assignments of individual variables. Thus the interpretation of x remains rigid, that is, intuitively for every "world" w ∈ **K** and for every "world" v ∈ **K** the interpretation picks out in the case of x the same object d in **D** at w and v.
- The interpretation of a predicate variable P may vary from world to world, that is, there are assignments in K that assign to P different extensions in the domain D.
- Given that there is no accessibility relation, □ expresses a universal S5 notion of necessity.

In Kripke (1963b) this picture is significantly changed. In particular, models with varying domains emerge, that is, each possible world come with its own set of individuals. In the first footnote, Kripke tells the readers that

The authors closest to the present theory appear to be Hintikka and Kanger. The present treatment of quantification, however, is unique as far as I know, although it derives some inspiration from acquaintance with the very different methods of Prior and Hintikka. (Kripke 1963b, 83)

What is this unique treatment of quantification? Essentially, it is obtained by imposing a quantificational structure on a set of possible worlds (with an accessibility relation). This happens by

- relativizing the range of a quantifier to a possible world; each possible world is endowed with its own domain
- providing a semantic value for free individual variables through the notion of assignment in such a way that the individuals assigned to the free variables may come from any of the world-relative domains
- relativizing the notion of satisfaction to a possible world and an assignment.

The starting point of the technical implementation is the notion of *model structure*. It is an ordered triple $\langle G, K, R \rangle$ where K is the set of possible worlds, G is the actual world such that $G \in K$, and R is an accessibility relation on K that Kripke interprets as follows:

```
If \mathbf{H}_1 and \mathbf{H}_2 are two worlds, \mathbf{H}_1\mathbf{R}\mathbf{H}_2 means intuitively that \mathbf{H}_2 is "possible relative" to \mathbf{H}_1; i.e., that every proposition true in \mathbf{H}_2 is possible in \mathbf{H}_1. (Kripke 1963b, 84)
```

Kripke notes that the reflexivity of \mathbf{R} is a natural requirement and mentions that one may impose additional requirements, corresponding to various axioms of modal logic. Given a model structure $\langle \mathbf{G}, \mathbf{K}, \mathbf{R} \rangle$ a model is a binary function ϕ which assigns to each atomic formula P and possible world \mathbf{H} in \mathbf{K} , a truth-value $\phi(P, \mathbf{H})$ which is \mathbf{T} or \mathbf{F} . Given a model, one can then assign by induction truth-values for complex propositional formulae. The clause for modal formulae is:

```
\phi(\Box A, \mathbf{H}) = \mathbf{T} iff \phi(A, \mathbf{H}') = \mathbf{T} for every \mathbf{H}' \in \mathbf{K} such that \mathbf{HRH}'; otherwise \phi(\Box A, \mathbf{H}) = \mathbf{F}. (Kripke 1963b, 84)
```

Intuitively: A is necessary in **H** if and only if A is true in every possible world **H**' related to **H**.

A quantified model structure is $\langle \mathbf{G}, \mathbf{K}, \mathbf{R} \rangle$ together with a domain function ψ that assigns to every possible world \mathbf{H} in \mathbf{K} its own world-relative domain $\psi(\mathbf{H})$, that is, the set of individuals existing in \mathbf{H} . We are told that:

Notice, of course, that $\psi(\mathbf{H})$ need not be the same set for different arguments \mathbf{H} , just as, intuitively, in worlds other than the real one, some actually existing individuals may be absent while new individuals, like Pegasus, may appear. (Kripke 1963b, 85)

Let U be the set of all individuals that exist in some world or another in K, that is:

$$\mathbf{U} = \bigcup_{\mathbf{H} \in \mathbf{K}} \psi(\mathbf{H})$$

U'' is the *n*-ary Cartesian product of U with itself. A quantificational model is now defined as a binary function $\phi(P'', \mathbf{H})$ where the second variable ranges over possible worlds in \mathbf{K} and the first variable over predicate symbols of the underlying language. When n = 0, P'' is a propositional letter and thus $\phi(P'', \mathbf{H})$ is either \mathbf{T} or \mathbf{F} . For $n \ge 1$, $\phi(P'', \mathbf{H})$ is a subset of \mathbf{U}'' , which is somewhat surprising as one would expect $\phi(P'', \mathbf{H})$ to be a subset of the world-relative domain of \mathbf{H} , that is, a subset of the set of individuals that exist in the possible world \mathbf{H} . Kripke defines inductively $\phi(A, \mathbf{H})$, the truth value of the formula A in the possible world \mathbf{H} relative to an assignment of individuals in \mathbf{U} to the free variables of A:

- (i)* The case of propositional variables has been taken care of above.
- (ii)* If A is $P''(x_1, ..., x_n)$ and $n \ge 1$, given an assignment of the individuals $a_1, ..., a_n$ from **U** to the variables $x_1, ..., x_n$, then $\phi(P'', \mathbf{H}) = \mathbf{T}$ if the n-tuple $(a_1, ..., a_n)$ belongs to $\phi(P'', \mathbf{H})$, otherwise $\phi(P'', \mathbf{H}) = \mathbf{F}$, relatively to the given assignment.

Given these assignments, Kripke defines inductively the assignments for complex formulae. The steps for common logical connectives and the modal operator are straightforward. For quantificational formulae we have:

(iii)* If we have a formula $A(x, y_1, ..., y_n)$ where x and $y_1, ..., y_n$ are the only free variables such that a truth-value $\phi(A(x, y_1, ..., y_n), \mathbf{H})$ has been defined for each possible assignment to the free

49

³³ Kripke explains his reasons for this decision in the footnote 1 on page 86.

variables $x, y_1, ..., y_n$, then we define $\phi(\forall x A(x, y_1, ..., y_n), \mathbf{H}) = \mathbf{T}$ relative to an assignment $a_1, ..., a_n$ of elements of \mathbf{U} to the free variables $y_1, ..., y_n$ if $\phi(A(x, y_1, ..., y_n), \mathbf{H}) = \mathbf{T}$ for every assignment of $b, a_1, ..., a_n$ to the free variables $x, y_1, ..., y_n$, where b is also an element of $\psi(\mathbf{H})$; otherwise $\phi(\forall x A(x, y_1, ..., y_n), \mathbf{H}) = \mathbf{F}$ (relatively to the given assignment).

The last restriction in (iii)* means that we quantify only over the individuals existing in **H**.

Kripke illustrates the above definitions by constructing counter-models to two famous candidates for laws of quantification theory in modal contexts, namely the Barcan formula

$$\forall x \Box Fx \rightarrow \Box \forall x Fx$$

and its converse

$$\Box \forall x Fx \rightarrow \forall x \Box Fx$$

For the first, consider a model structure $\langle \mathbf{G}, \mathbf{K}, \mathbf{R} \rangle$ where \mathbf{K} consists of two worlds, the actual world \mathbf{G} and a second world \mathbf{H} . The accessibility relation \mathbf{R} is symmetric, reflexive, and transitive. The quantificational model structure on $\langle \mathbf{G}, \mathbf{K}, \mathbf{R} \rangle$ is formed by endowing each possible world with its own domain. In the case of the Barcan formula take: $\psi(\mathbf{G}) = \{a\}$ and $\psi(\mathbf{H}) = \{a, b\}$. Finally, to obtain a model, we have to define an extension of the predicate symbol F in each possible world. Following Kripke, let: $\varphi(F, \mathbf{G}) = \{a\}$ and $\varphi(F, \mathbf{H}) = \{a\}$. Now $\Box Fx$ is true in \mathbf{G} under the assignment of a to x. Since a is the only individual in $\mathbf{G}, \forall x \Box Fx$ is also true in \mathbf{G} , that is, $\varphi(\forall x \Box Fx, \mathbf{G}) = \mathbf{T}$. But $\forall x Fx$ is false in \mathbf{H} and thus $\Box \forall x Fx$ is false in \mathbf{G} , that is, $\varphi(\Box \forall x Fx, \mathbf{G}) = \mathbf{F}$. Thus the antecedent of the Barcan formula is true in \mathbf{G} and its consequent is false in \mathbf{G} . By the assumed semantical clause for implication, the Barcan formula itself is false in \mathbf{G} .

For the converse Barcan formula, take: $\psi(\mathbf{G}) = \{a, b\}$ and $\psi(\mathbf{H}) = \{a\}$ and let: $\varphi(F, \mathbf{G}) = \{a, b\}$ and $\varphi(F, \mathbf{H}) = \{a\}$. Now $\forall x F x$ is true in both worlds, that is, $\varphi(\forall x F x, \mathbf{G}) = \mathbf{T}$ and $\varphi(\forall x F x, \mathbf{H}) = \mathbf{T}$, given that any assignment of an element of $\psi(\mathbf{G})$ to x is a member of $\varphi(F, \mathbf{G})$ and similarly for $\varphi(\forall x F x, \mathbf{H}) = \mathbf{T}$. Thus $\Box \forall x F x$ is true in \mathbf{G} . But F x is false in \mathbf{H} when b is assigned to x, so $\varphi(\Box F x, \mathbf{G}) = \mathbf{F}$. Hence $\forall x \Box F x$ is false in \mathbf{G} , that is, $\varphi(\forall x \Box F x, \mathbf{G}) = \mathbf{F}$. By the clause for implication the converse Barcan formula is false in \mathbf{G} .

³⁴ Kripke ignores some of the subtleties of Tarskian truth definitions: for instance, in the present paper he does not address the relationship of satisfaction and truth, and here he ignores the fact that $\forall x \Box Fx$ and $\forall x Fx$ are formulae with no free variables, that is, sentences, and his exposition here does not contain semantical clauses for sentences.

Thus we have two counter-models to establish that the Barcan formulae are not *universally valid*. In both cases the individual b is crucial to the counter-example: in the first case b comes into existence when we move from world \mathbf{G} to \mathbf{H} and in the second case b drops out of existence when we move from \mathbf{G} to \mathbf{H} . If coming into and dropping out of existence are prevented by, for instance, defining a common constant domain as in Kripke (1959b) then these counter-examples cannot be constructed and both Barcan formulae would remain valid.

Kripke (1963b, 84) states that the proof he had given earlier in (Kripke 1963a) for propositional modal logics may be extended to prove the completeness of several quantified systems in relation to this new semantics. These systems are *inclusive* in Quine's sense (empty domain is *included*) and only closed formulae without free variables are asserted. The axioms of quantified system Kripke calls "M" are defined as the closures of

- (0) Truth-functional tautologies
- (1) $\square A \rightarrow A$
- $(2) \quad \Box(A \to B) \to (\Box A \to \Box B)$
- (3) $A \rightarrow \forall x A$, provided that x is not free in A
- $(4) \quad \forall x (A \to B) \to (\forall x A \to \forall x B)$
- (5) $\forall y (\forall x Ax \rightarrow Ay)$

The only rule of inference is *modus ponens*, while necessitation may be obtained as a derived rule. The quantified extensions of S4, S5, and B may be obtained by adding the characteristic modal axiom schemas respectively. I shall not go into these details for their relevance is limited to the semantic comparison at hand.

I return to the Barcan formulae in sections 2.8 and 2.9. As we shall see, the formulae are invalid in Hintikka's deontic and alethic logics, but in epistemic logic the converse Barcan formula is valid. Now I move on to survey Hintikka's work.

2.4 HINTIKKA'S MODEL SETS

Hintikka (1955) introduced model sets as a new tool in logical semantics, and constructed a new proof of the completeness of first-order logic. Hintikka thought that his model sets provided a self-sufficient method that overcomes "the dualism of form and content", that is, a method that avoids the shortcomings of syntactical axiomatic methods and semantical "set-theoretic" approaches.³⁵ As a consequence,

³⁵ According to Hintikka, the standard methods were, among other things, circular:

^[...] the dichotomy of form and content appears as a dualism of two different approaches. Existing side by side, they may be characterized as the semantical and the syntactical approach. [...] The purely formal approach does not [...] explain why one particular system of axioms and transformation rules is chosen among the multitude of choices [...] And the so-called completeness proofs which do explain the choice fall back on set-theoretic concepts if not on set-theoretic methods. (Hintikka 1955, 18)

Hintikka largely neglected both proof theory and model theory in his early writings and studied and relied on model sets instead as providing a comprehensive method for his purposes.

A model set is a set of sentences in the relevant logical language that constitutes a partial description of a possible state of affairs. One starts with a first-order language L. A model set μ is any set of sentences of L that satisfies some closure conditions derived from the intuitive meanings of logical connectives. The conditions are given in several of Hintikka's early writings (the exact formulations of conditions as well as their names vary slightly from one source to another):

- (i) For any atomic sentence A, not both $A \in \mu$ and $\neg A \in \mu$.
- (ii) If $A \wedge B \in \mu$, then both $A \in \mu$ and $B \in \mu$.
- (iii) If $A \vee B \in \mu$, then either $A \in \mu$ or $B \in \mu$.
- (iv) If $\neg \neg A \in \mu$, then $A \in \mu$.
- (v) If $\neg (A \land B) \in \mu$, then $\neg A \in \mu$ or $\neg B \in \mu$.
- (vi) If $\neg (A \lor B) \in \mu$, then $\neg A \in \mu$ and $\neg B \in \mu$.

Further conditions are introduced for quantified sentences:

- (C.E) If $\exists x A \in \mu$, then $A(x/b) \in \mu$ for at least one constant b of L.
- (C.U) If $\forall x A \in \mu$, and if b occurs in at least one member of μ , then $A(x/b) \in \mu$.
- (C.¬E) If $\neg \exists x A \in \mu$, then $\forall x \neg A \in \mu$.
- (C. \neg U) If $\neg \forall x A \in \mu$, then $\exists x \neg A \in \mu$.

Identity requires additional rules:

(C.=) If A is an atomic sentence or its negation, and $A \in \mu$ and if B is exactly like A except that a and b have been interchanged in one or several places, then $B \in \mu$.

(C.self \neq) For no constant b: $b \neq b \in \mu$.

Sometimes Hintikka prefers the following rule to (C.self≠):

(C.self=) If b occurs in the sentences of μ , then $b = b \in \mu$.

The purpose of studying the notion of model set is expressed, for instance, in the following passage:

The basic notion of a semantical theory is normally the notion of truth. In so far as we are not interested in truth under *some particular* interpretation of logical formulae but rather in the question whether there are *any* interpretations which make a given set of formulae true

(in short, if we are not interested in any one interpretation more than in the others), the basic concept of a semantical theory may also be chosen to be that of *satisfiability*. [...] If the negation of a formula *f* is not satisfiable, *f* is said to be *valid*.³⁶ (Hintikka 1961, 119)

Hintikka defined the notion of satisfiability by reference to the Carnapian notion of *state-description*:

A set of formulae [i.e. sentences] λ is satisfiable if and only if there is a statedescription in which all the members of λ hold. (Hintikka 1961, 119)

For a single sentence \mathcal{A} , it is said that \mathcal{A} is satisfiable if and only if its singleton set, $\{\mathcal{A}\}$, is satisfiable. Thus a sentence is satisfiable if and only if \mathcal{A} holds in a state-description. A state-description, a precursor of the notion of model set, is a set of sentences that satisfies the following conditions:

- (C.1) If A is an atomic sentence (or an identity), then not both $A \in \mu$ and $\neg A \in \mu$.
- (C.2) If A is an atomic sentence (or an identity), then either $A \in \mu$ or $\neg A \in \mu$.
- (C.3) If A is an atomic sentence (or an identity) or the negation of an atomic sentence (identity), and if $A \in \mu$ and $a = b \in \mu$ and if B is exactly like A except that a and b have been interchanged in one or several places, then $B \in \mu$.
- (C.4) Not $\neg (b = b) \in \mu$.

Thus, a state-description is essentially a set of *atomic sentences* or their negations. In order to understand the above definition of satisfiability, we still need to understand what it is for *all the members of* λ *to hold in a state-description*. One way to proceed, following Hintikka (1961), is to give necessary and sufficient conditions for a set of sentences μ to be the set of *all* sentences which hold in a state-description, including complex sentences containing \wedge , \vee , $\exists x$, and $\forall x$. The set of conditions includes, in addition to (C.1) - (C.4), the following:

- (C.5) If $A \wedge B \in \mu$, then both $A \in \mu$ and $B \in \mu$.
- (C.6) If both $A \in \mu$ and $B \in \mu$, then $A \wedge B \in \mu$.
- (C.7) If $A \vee B \in \mu$, then either $A \in \mu$ or $B \in \mu$.
- (C.8) If either $A \in \mu$ or $B \in \mu$ and all the individual constants occurring in $A \vee B$ occur in the other sentences of μ , then $A \vee B \in \mu$.
- (C.9) If $\exists x A \in \mu$, then $A(x/b) \in \mu$ for at least one constant b.

³⁶ Hintikka uses the term *formula* but strictly taken he means *sentence*. A model set is a set of sentences and the model set conditions apply only to formulae with no free variables, that is, to sentences.

- (C.10) If $A(x/b) \in \mu$ for at least one constant b, then $\exists x A \in \mu$.
- (C.11) If $\forall x A \in \mu$, and if b occurs in at least one member of μ , then $A(x/b) \in \mu$.
- (C.12) If $A(x/b) \in \mu$ for every individual constant b which occurs in the sentences of μ , then $\forall x A \in \mu$.

Thus conditions (C.1) – (C.4) make sure that μ is a Carnapian state-description, and the other conditions constitute a recursive definition of what it is for a complex sentence to hold in a state-description. The clauses for negation are missing, because it is assumed that the negation occurs only in front of an atomic sentence. But if this assumption were dropped, they could be easily added, for instance:

- (C.13) If $\neg (A \land B) \in \mu$, then either $\neg A \in \mu$ or $\neg B \in \mu$.
- (C.14) If either $\neg A \in \mu$ or $\neg B \in \mu$, then $\neg (A \land B) \in \mu$.

Now Hintikka is in a position to restate the above definition of satisfiability of a set of sentences and reformulate it as follows:

```
a set of formulae [i.e. sentences] is satisfiable if and only if it can be imbedded in a set which satisfies conditions (C.1) – (C.12). (Hintikka 1961, 121)
```

One of Hintikka's insights in his early work was that many of the conditions above are redundant: for instance, all the right-to-left conditions are redundant for his purpose. He ends up with only the left-to-right conditions (C.1), (C.3), (C.4), (C.5), (C.7), (C.9), and (C.11), and calls any set that satisfies them a *model set*. Together with similar left-to-right conditions for negation, the conditions constitute his definition of a model set. Hintikka is then able to prove that a set of sentences is satisfiable if and only if it can be embedded into a model set. I do not go into the execution of this proof, but I quote a paragraph in which Hintikka expresses this result informally:

The result may perhaps be expressed intuitively by saying that a model set is the formal counterpart of a possible state of affairs (of a 'possible world'). (It is, however, large enough a description to make sure that the state of affairs in question is really possible.) For it is natural to say that a set of sentences is satisfiable if and only if it can be imbedded in a (partial or exhaustive) description of possible states of affairs; and this is just what we demonstrated if model sets are interpreted as such descriptions. (Hintikka 1961, 122)

Through the notion of satisfiability, Hintikka was able to define other important notions. A sentence A is *contradictory* if it is not satisfiable, and A is *valid* if and only if $\neg A$ is not satisfiable, that is, $\neg A$ is contradictory.

In addition to the notion of model set, Hintikka also acknowledged, at this point, a notion of *model* similar to the one we encountered in the previous section, that is, a model determined by the domain of individuals, interpretation of individual constants by the individuals in the domain, and the interpretation of predicate symbols by the *n*-tuples of individuals. Hintikka emphasized that the notions of model and model set are distinct, but as the choice of the name "model set" suggests one may recover from every model set a model (in fact many models) in which all the sentences comprising a model set are true under some "suitable interpretation" (Hintikka 1955, 26). Regarding the notion of true in a model, Hintikka gave similar truth-definitions, typical to model theory, which we saw in the previous section. For instance, for atomic sentences and negation Hintikka gave the clauses

$$P(a_1, ..., a_n)$$
 is true iff $(a_1, ..., a_n)$ belongs to the extension of $P \neg P(a_1, ..., a_n)$ is true iff $P(a_1, ..., a_n)$ is not true

With similar truth-definitions for all logical connectives and quantified sentences in place, Hintikka defined, given a set of sentences λ , a model of λ in which all the sentences in λ are true. If such a model exists, then λ is said to be satisfiable. Now it follows that every set of sentences satisfiable in this sense also fulfills Hintikka's model set conditions, as these conditions simply paraphrase the central content of truth-definitions in model set terms. Hence model sets – and the notion of satisfiability in particular – were backed up by proper model-theoretical definitions. By establishing, first, that a given sentence is satisfiable just in case it can be embedded into a model set and, second, that from every model set a model may be recovered, Hintikka secured a link between model sets and models. Thereby the satisfiability of a sentence such as Pa is not merely a matter of fulfilling (or not violating) syntactic model set conditions (C.1) – (C.5) but also a matter of there being an individual i (in the domain of individuals) assigned as the semantic value of the individual constant a and there being a set of individuals assigned as the semantic value of the predicate a and a belonging to that set.

Compared to the models discussed by Kripke, Hintikka's models were, first, designed for non-modal languages and thereby came without possible worlds or related entities. Second, Hintikka's models were notably linguistic: the domain of individuals consisted of individual constants and the interpretation of each constant was the constant itself, while the interpretation of a predicate was naturally a set of *n*-tuples of individual constants. Quantification was *substitutional*, governed by clauses such as

 $\forall x A$ is true iff A(a/x) is true for every individual constant a.

Kripke's clauses for universal quantifier (iv) and (iii)* in the previous section are very different as they display an *objectual* interpretation of quantifiers (see more on this below in section 2.8). When Hintikka moved on to study modal languages he dropped the notions of model and truth altogether, as we shall see in the next section.

2.5 DEONTIC MODAL LOGIC

"Quantifiers in Deontic Logic" (1957b) was Hintikka's first contribution to modal semantics. Some parts of it appeared later in Hintikka (1971), which was a more comprehensive exposition of problems in deontic logic. In Hintikka (1957b) the notion of satisfiability was extended to sets of sentences that contain modal operators. The key insight concerning modality was that the satisfiability of modal sentences may be defined in terms of sets of model sets, that is, in terms of *model systems*. Intuitively a model set is a linguistic counterpart of a (partial) possible world, and a model system a linguistic counterpart of modal space. As the title suggests, the discussion concerned reasoning involving deontic concepts, and hence the relevant modal notions were obligation and permission. The alternativeness relation, in turn, was called "copermissibility".

Hintikka (1957b) considers a deontic language that contains "free individual variables" a, b, c, \ldots ranging over *individual acts*, n-adic predicate variables A^n , B^n , C^n , ... ranging over *properties of acts*, the common logical connectives (no identity), quantifiers with bound variables x, y, z, \ldots and deontic modal operators O (obligation), P (permission), and F (forbiddance). Hintikka did not adopt von Wright's (1951a, 1951b) propositional language as the language of deontic logic, that is, he did not accept the representations OA, PA, or FA, where the operators are applied to propositional formulae and not directly to properties of individual acts. Neither did he adopt Prior's (1955) representation employing letters a, b, c, \ldots for properties of acts, to be read "an act of the sort a is done". Prior's formalism consisted of formulae such as Pa, $O \neg b$, and $P(a \land b)$ to mean "an act of the sort a is permitted", and so forth. Instead, Hintikka used quantifier/variable notation to denote individual acts (performed in a particular situation). Thus Prior's $(a \land b)$ was rendered as

 $\exists x Ax \land \exists x Bx$

The deontic operators are applied to first-order sentences. Hintikka starts with the notion of forbiddance. To say that acts of a kind \mathcal{A} are forbidden

is to say that no act under consideration may be of this kind. In other words, every act under consideration ought to be an omission of A,

i.e., every act ought to be an instance of the negation $\neg A$ of A. (Hintikka 1957b, 5)

Hence Hintikka expresses von Wright's FA, the forbiddance of A, by $\forall x O \neg Ax$ or by $O \forall x \neg Ax$. If so, then "A is obligatory" would come out as $\forall x O \neg \neg Ax$, that is, $\forall x O Ax$. Hintikka states that while such obligations exist they are not primary. When one ought to pay one's taxes, it does not mean that one's every act ought to be a taxpaying act (Hintikka 1957b, 6). Rather, the act of paying taxes should be among the things one does. The logical form of these obligations is $O \exists x Ax$. Further, an act is permitted if in every particular situation one is allowed to perform an act of this kind, $\forall x PAx$.

The main goal of the paper is to formulate the notion of satisfiability for sentences with deontic operators. Hintikka (1957b, 10) writes that his treatment follows not just his earlier treatment of quantification theory but also his new general theory of modal logic developed in an unpublished manuscript (this manuscript is presumably the one I referred to above as the *Boston Manuscript*). Hintikka is explicit that he will focus on aspects that are important for philosophical applications and not on proofs. Perhaps his most important insight is that the satisfiability of a set of sentences involving modal notions forces us to consider sets of model sets, that is, model systems:

Let us first consider formulae of the form [PA]. What do we mean by saying that [A] is permitted? Obviously, the content of this statement cannot be exhausted by speaking of what actually takes place. When speaking of permissions, we are not really speaking of the actual state of affairs at all. Rather, we are speaking of something that *could* have taken place [...] This means that we cannot formulate rules for a formula of the form [PA] occurring in μ in terms of μ , for μ was thought of as being concerned with the actual state of affairs. Rather, we must consider, in addition to μ , another set μ^* related to μ in a certain way. This relation will be expressed by saying that μ^* is *copermissible with* μ . We may think of μ^* as being concerned with the (imagined) state of affairs in which [A] was supposed to take place. (Hintikka 1957b, 11)

Thus the definition of the satisfiability of modal sentences requires a set of model sets, a model system, together with an alternativeness relation between model sets, which in the context of deontic logic is called copermissibility. Hintikka inquires into the question of what conditions model systems must be subject to. One of them, motivated in the preceding quote is:

(C.5) If $PA \in \mu$, then there is at least one model set μ^* copermissible with μ such that $A \in \mu^*$.

Hintikka requires even more: if μ^* is copermissible with μ , then everything one in fact ought to do in μ is thought of as being done in μ^* . That is, Hintikka also adopts the condition

(C.6)† If $OA \in \mu$, and if μ^* is copermissible with μ , then $A \in \mu^*$.

And he also points out that actually existing obligations should also hold in copermissible worlds:

(C.6) If $OA \in \mu$, and if μ^* is copermissible with μ , then $OA \in \mu^*$.

There is one more condition on obligations, which derives from the fact that copermissible worlds are thought of as deontically perfect worlds where all obligations are fulfilled. That is, if a world μ^* is copermissible with a given world μ and A is obligatory in μ^* , then A is the case in μ^* :

(C.7) If $OA \in \mu^*$ and if μ^* is copermissible with some other set μ of sentences, then $A \in \mu^*$.

Hintikka (1957b) studies a language without identity, and all the sentences are in the negation normal form, that is, the negation sign occurs only in front of atomic sentences. For this reason the earlier conditions on model sets reduce to the following:

- (C.0) If $A \in \mu$, then not $\neg A \in \mu$.
- (C.1) If $A \wedge B \in \mu$, then $A \in \mu$ and $B \in \mu$.
- (C.2) If $A \vee B \in \mu$, then $A \in \mu$ or $B \in \mu$.
- (C.3) If $\exists x A \in \mu$, then $A(x/a) \in \mu$ for some free individual variable a.
- (C.4) If $\forall x A \in \mu$, then $A(x/b) \in \mu$ for every free individual variable b occurring in the sentences of μ .

Now we are almost in a position to state the definition of satisfiability for deontic sentences. The one missing piece is the notion of model system: a *model system* is a pair $\langle \Omega, R \rangle$ where Ω is a set of model sets such that each member of Ω is a set of sentences that satisfies (C.0) – (C.4) and R is an accessibility relation on Ω (Hintikka 1961, 123).

And now to the notion of satisfiability. The idea is the same as above in the case of model sets for first-order languages: deontic sentences are satisfiable when they are embeddable into a model set belonging to a model system, and valid when their negations are not satisfiable. Hintikka defines a finite set of sentences as satisfiable if and only if it can be embedded into a model system $\langle \Omega, R \rangle$ where R fulfills the conditions (C.5), (C.6), and (C.7) on the copermissibility relation:

A set λ of formulae [i.e. sentences] is satisfiable if and only if there is a model system $\langle \Omega, R \rangle$ such that $\lambda \subseteq \mu$ for some member μ of Ω . (Hintikka 1961, 123)

Hintikka defines satisfiability exclusively in terms of *model system* and related concepts. The notions of model and truth as encountered in Hintikka (1955) and Kripke (1959b) & (1963b) are absent.

As Hintikka (1957b, 13) points out, one of the strengths of this approach is that this definition also gives a *method* to show that a finite set of sentences λ is satisfiable: Produce a model system $\langle \Omega, R \rangle$ such that

- (i) $\lambda \subseteq \mu$ for some member μ of Ω , and
- (ii) ensure that the conditions (C.0) (C.7) are fulfilled.

As A is defined to be valid if and only if $\{\neg A\}$ is not satisfiable, Hintikka states that we also need a procedure to show that a finite set λ of sentences is *not satisfiable* (Hintikka 1957b, 13). He introduced a different set of rules, E-rules, for this purpose but he utilized them only in his deontic logic. When he moved on to alethic and epistemic logics he dropped the E-rules and exclusively used his C-rules in the proofs.³⁷ In any case, here is Hintikka's E-rule procedure.

We start with $\{\lambda\}$ as our initial set of sets of sentences and the empty copermissibility relation and try to build up a model system by adjoining new sentences to λ and new sets of sentences copermissible in an appropriate way. Every adjunction is supposed to remove one violation of the closure conditions mentioned above. Following this procedure we build at each stage a set of sets of sentences Ω' related by a co-permissibility relation R'. Hintikka (1957b, 13) calls such sets *provisionally copermissible*. Thus suppose Ω' is the set of sets of sentences built up so far. We apply the following instructions to govern the introduction of new sentences and sets:

- (E.11) If $\mu \in \Omega'$ and $A \wedge B \in \mu$ but not $A \in \mu$, then we may adjoin A to μ .
- (E.12) If $\mu \in \Omega'$ and $A \wedge B \in \mu$ but not $B \in \mu$, then we may adjoin B to μ .
- (E.2) If $\mu \in \Omega'$ and $A \vee B \in \mu$ but neither $A \in \mu$ nor $B \in \mu$ then we may adjoin either A or B to μ .
- (E.3) If $\mu \in \Omega'$ and $\exists x.A \in \mu$, but not $A(x/b) \in \mu$ for any free individual variable b, then we may adjoin A(x/a) to μ provided that a is a free individual variable not occurring in the sentences of any member of Ω' .
- (E.4) If $\mu \in \Omega'$ and $\forall x A \in \mu$ but not $A(x/b) \in \mu$ for some free individual variable b occurring in the sentences of μ , then we may adjoin A(x/b) to μ .

³⁷ E-rules may be construed from the corresponding C-rules.

- (E.5) If $\mu \in \Omega'$ and $PA \in \mu$ but not $A \in \mu^*$ for any μ^* provisionally copermissible with μ , then we may adjoin $\{A\}$ as a new model set to Ω' and stipulate that it is provisionally copermissible with μ .³⁸
- (E.6) If μ , $\mu^* \in \Omega'$ and $OA \in \mu$ but not $OA \in \mu^*$ and if μ^* is provisionally copermissible with μ , then we may adjoin OA to μ^* .
- (E.7) If $\mu^* \in \Omega'$ and $OA \in \mu^*$ but not $A \in \mu^*$, then we may adjoin A to μ^* provided that μ^* is provisionally copermissible with some $\mu \in \Omega'$.

Here, then, is the procedure to show that a set of sentences λ is not satisfiable, using the above rules: We start from a set of sets of sentences containing only λ as its only member (and the empty copermissibility relation) and we apply the rules (E.2) – (E.7), (E.11), and (E.12) repeatedly to this set of sets. Every time we apply the rule (E.5) we add a new set with the appropriate accessibility (copermissibility) relation. And every time we apply the rule (E.2) to a given set, the result consists of two sets of sets (each with its own copermissibility relation). Each of them is a basis for further applications of the rules. We try to arrange the rules in such a way that we eventually end up with a set of sets in which one of the members violates the rule (C.0), that is, one model set in the model system contains a sentence together with its negation. If this happens in all the alternative ways in which the E-rules can be applied

then it is easy to show that the set λ from which we started has no model set, i.e. that it is contradictory. This is our basic *soundness theorem*. Although it is easy to establish, we shall not try to prove it here [...] It is also known (although we have not proved it here) that if λ is contradictory, then the situation of this kind can somehow be reached. This is the central *completeness result* of our deontic logic. (Hintikka 1971, 77)

This paragraph raises many questions. What exactly do *soundness* and *completeness* mean in Hintikka's syntactically oriented setting? Why call the latter "the central completeness result" when there is no proof of it? In the same vein, what does he mean by "it is also known" if, again, there is no proof to provide that knowledge? And finally, why does he state "it is easy to establish"? The quote is from 1971, fourteen years after the initial publication of the deontic study under scrutiny. If it were easy, surely he or somebody else would have elaborated on these concepts and established whatever meta-logical properties they refer to by providing proofs. But the fact is that no proofs were published.

Let us put together what Hintikka calls the "soundness theorem" and the "completeness result" of his deontic logic to get:

³⁸ Hintikka (1957b, 14) assumes that all relations of provisional copermissibility arise from the application of this rule, (E.5).

• λ is satisfiable if and only if λ has a model set

or, equivalently

• λ is not satisfiable if and only if λ has no model set

Although he never proved it, Hintikka used this alleged result for the next ten years in all his writings on modal logics, and after that in all the writings in which he chose to employ model sets: Whenever he wanted to show that a sentence A is valid, he used the definition A is valid if and only if $\neg A$ is not satisfiable and then established that $\neg A$ is not satisfiable by appealing to the above "theorem" and by showing that $\neg A$ has no model set. Hence for a single sentence we have

• A is valid if and only if $\neg A$ has no model set

Whereas Kripke's (1959b) completeness proof connects the model-theoretically defined semantic concept of validity to the syntactic notion of provability defined in terms of tableau rules, Hintikka's alleged result connects the notion of validity and the notion of not having a model set, which are both defined in terms of his model set conditions. I shall return to this shortcoming in section 2.8.

To illustrate Hintikka's method, I choose the following sentence that Hintikka considers in deontic, alethic, and epistemic contexts. The deontic variant of this formula appears in (Hintikka 1957b, 22) as

$$(49) \qquad \exists x O A x \to O \exists x A x$$

Hintikka does not present a formal argument for the validity of (49) but it is easy to build one. Let us establish that (49) is valid in the deontic system at hand, that is, show that its negation is not satisfiable, or that "its antecedent cannot be satisfied together with the negation of the consequent" (Hintikka 1957b, 21). Suppose, then, that there is a model set μ in a model system Ω such that it contains the negation of (49) or a "counter-assumption" of (49), as Hintikka often puts it.

1.
$$\exists x O A x \in \mu$$

2.
$$\neg O\exists x Ax \in \mu$$

By a series of equivalent transformations on (2), we get

3.
$$P \forall x \neg Ax \in \mu$$

which, by (E.5), implies

4.
$$\forall x \neg Ax \in \mu^*$$

where μ^* is copermissible with μ . Now, focusing on the antecedent, from (1) and (E.3) we get

5.
$$OAa \in \mu$$

and by applying (E.6) to (5) we get

6.
$$OAa \in \mu^*$$

which together with (E.7) gives

7.
$$Aa \in \mu^*$$

Now, step (4) establishes that $\forall x \neg Ax$ belongs to μ^* . Since *a* occurs in the sentences of μ^* – as steps (5) and (6) indicate – we may instantiate $\forall x \neg Ax$ with the individual variable *a* in accordance with the rule (E.4) to get

8.
$$\neg Aa \in \mu^*$$

(7) and (8) are contradictory, and thus μ^* violates (C.0). The negation of (49) does not have a model set, that is, it is not satisfiable, and hence (49) is valid by Hintikka's definition. Hintikka, however, finds the validity of (49) intuitively unacceptable:

If there is, under the actual course of events, an act that ought to be an instance of forgiving a trespass, it clearly does not follow that there ought to be, under any deontically perfect course of events, an act of forgiving – and hence, presumably, also another earlier act of trespassing. (Hintikka 1971, 102)

According to Hintikka's intuitions $\exists x OAx$ does not imply $O\exists xAx$ because an actually performed obligatory act a may not occur in the deontic alternatives. Hence, in possible worlds jargon, individual acts may drop out of existence when we move from the actual world μ to a possible world μ^* . Of course, officially μ and μ^* are not "possible worlds" but model sets, and a is not an individual act taking place in a possible world but an individual variable, a linguistic entity, occurring in a sentence in a model set. Nevertheless, Hintikka freely invoked such considerations to back up his theoretical decisions. In any case Hintikka is here tacitly addressing a question relevant to the validity of the converse Barcan formula, namely whether individuals may drop out of existence when we move from one possible world to another. Hintikka's answer is affirmative, which means that at least some of the deontic sentences obtainable from a counterpart of the converse Barcan schema are invalid in Hintikka's deontic logic.

How about our proof of (49), that is, $\exists x OAx \rightarrow O\exists xAx$? The problem with the above derivation is condition (E.6), which assumes that things do not drop out of existence when we move from the actual world to one of the alternatives (Hintikka 1957b, 22). We applied it in the step from (5) to (6): we had OAa in μ and then by (E.6) we established OAa also in μ^* assuming, among other things, that the individual act a in μ is also performed in μ^* . This is an unwarranted "existential presupposition" because

for any given act occurring under certain circumstances we can imagine a course of events under which this act does not occur. (Hintikka 1957b, 23)

To block the derivation we must replace (E.6) by

(E.6)* If μ , $\mu^* \in \Omega'$ and $OA \in \mu$ but not $OA \in \mu^*$ and if μ^* is provisionally copermissible with μ and all the free individual variables in A occur in some sentence of μ^* , then we may adjoin OA to μ^* . (Hintikka 1971, 101)

This modification prevents the step from (5) to (6) and, in general, the resulting semantics allows individuals to drop out of existence when we move from one possible world to another. Still, Hintikka adds the following reservations:

This underlying idea may be challenged. Are there not, it may be asked, acts so indispensable that they must be assumed to be performed under any circumstances? And are there not in any case acts which are hypothetically necessary in the sense that they are made necessary by certain other acts? (Hintikka 1957b, 23)

Hintikka thought that these questions must be answered within a logic for necessity and possibility, which fell outside the scope of his paper. He returned to these matters in "Modality and Quantification" (1961).

Hintikka nevertheless commented on the question of whether individuals may come into existence when moving from one possible world to another. He showed that the deontic counterpart of the Barcan formula, namely

$$(41) P\exists xAx \to \exists xPAx$$

is not valid in his deontic logic. However, (41) may be validated if the quantifier rules (C.4) and (E.4) are modified by replacing the condition

 $A(x/b) \in \mu$ for every/some free individual variable b occurring in the formulae of μ

with the more liberal

 $A(x/b) \in \mu$ for every/some free individual variable *b* occurring in some member of the model system (Hintikka 1957b, 21).

This modification amounts to an assumption that if b occurs in any model set then it occurs in all model sets, that is, in possible worlds terms, if b exists in one possible world then it exists in all alternative worlds. Let us now establish that with these assumptions at hand, the Barcan formula is valid. As above, suppose there is a model set μ in a model system Ω such that

1.	$P\exists xAx$	$\in \mu$	
2.	$\forall x O \neg A x$	$\in \mu$	counter-assumption
3.	$\exists x A x$	$\in \mu^*$	from 1 by (E.5)
4.	Aa	$\in \mu^*$	from 3 by (E.3)
5.	$O \neg Aa$	$\in \mu$	from 2 by modified (E.4)
6.	$O \neg Aa$	$\in \mu^*$	from 5 by (E.6)
7.	$\neg Aa$	$\in \mu^*$	from 6 by (E.7)

(4) and (7) violate (C.0) and hence (41) is valid. Note that the modified rule (E.4) allows us to instantiate $\forall x O \neg Ax$ with a, which does not occur in the sentences of μ . The original (E.4) does not justify this instantiation but the modified (E.4) does. Hintikka, however, finds (41) unintuitive because an act of the kind A may be permitted even though no actual act of this kind is performed (Hintikka 1957b, 21). So he rejects the proposed liberalization of the condition (E.4) in the context of deontic notions. By modifying operator/quantifier rules, Hintikka is nevertheless in a position to manipulate the behavior of his variables in different model sets, or in possible worlds jargon, in a position to control the domains of possible worlds by invoking increasing or decreasing domains if needed.

Hintikka (1957b) introduced the main method that Hintikka followed for many years in his studies of various modal logics: model systems for establishing validities and intuitive restrictions on quantifier/operator rules and on the various assumptions that individuals existing according to one model set also exist according to other model sets. Hintikka considered these *semantical* or *quasi-semantical* methods and found them more transparent and natural than conventional proof-theoretical methods. Hintikka concluded:

It turns out that an axiomatic and deductive system of quantified deontic logic cannot be built without limiting the applicability of the [quantifier/operator] rules [...] The trouble with a conventional axiomatic approach is that it makes extraordinarily hard to see the nature and the rationale of the requisite modifications of the different rules. (Hintikka 1957b, 23)

However, from a model-theoretic point of view the model set approach is not as transparent as Hintikka declares. Semantical questions concerning individuals, truth, and possible worlds are impossible to address directly within that framework. For instance, we are told that $\exists x OAx \rightarrow \neg O\exists xAx$ is satisfiable. Let us say that the antecedent is true in the actual world under the assignment of the individual a to x. The falsity of the consequent requires that there are accessible alternative worlds where a does not exist. In such worlds, is Ax under the assignment of a to x true, false, or without truth-value? Further, does OA mean that A is true in the alternative worlds or that A is not false, that is, that A is either true or without a truth-value? (E.6)* suggests that the answer to first question is either "false" or "without truth-value", and the satisfiability $\exists x OAx \rightarrow \neg O\exists x Ax$ suggests that it is "true" or "without truth-value". If the answer is "without truth-value" then OA must mean that A is not false in the alternative worlds.

Due to its syntactical orientation Hintikka's model set approach is not suitable for direct assessment of semantical questions concerning reference and truth. But it is concise and flexible for many applications. By laying out a few conditions and rules Hintikka had at his disposal a comprehensive method for developing many kinds of modal logics. Hintikka's more famous subsequent studies on alethic and epistemic modal logic add nothing to the core methodology introduced in Hintikka (1957b).

2.6 ALETHIC MODAL LOGIC

In "Modality and Quantification" (1961) Hintikka studied the operators □ (necessity) and ⋄ (possibility), and addressed questions concerning identity in modal contexts. There are two different versions of the paper, the original published in *Theoria* 27 and the expanded version included as a chapter in *Models for Modalities* (1969). From a technical point of view these expositions add nothing to Hintikka (1957b).

Hintikka introduced (C.M*) as the counterpart of the deontic condition (C.5) to govern the possibility operator \diamondsuit and the accessibility or alternativeness relation between model sets:

(C.M*) If $\Diamond A \in \mu \in \Omega$, then there is in Ω at least one alternative μ^* to μ such that $A \in \mu^*$. (Hintikka 1961, 123)

For the necessity operator □ at least the following condition must hold because "we have to require that what is said to happen necessarily happens actually" (Hintikka 1961, 123):

(C.N) If $\Box A \in \mu \in \Omega$, then $A \in \mu$

Further, Hintikka contemplated the following condition

(C.N⁺) If $\Box A \in \mu \in \Omega$, then there is in Ω at least one alternative μ^* to μ such that $A \in \mu^*$.

but noticed that if applied unrestrictedly it introduces unacceptable existential assumptions:

The presence of a free individual variable in the formulae of μ , we may thus say, is the formal counterpart to the existence of its value in the state of affairs described by μ .

From this it follows that when a formula f is transferred from a model set μ to one of its alternatives – say ν – we have to heed the free individual variables f contains. If one of them does not occur in the other formulae of ν , then the adjunction of f to ν is legitimate only if the relevant values of this free individual variable are assumed to exist not only in the state of affairs described by μ but also in that described by ν . In general, this assumption cannot be made. Individuals which de facto exist may possibly fail to do so. (Hintikka 1961, 124 – 125)

Hence Hintikka rejected these existential assumptions for the same reasons he rejected them in the context of deontic logic and, consequently, restricted the condition (C.N⁺) in the same way that he restricted the deontic conditions (C.6) and (E.6):

(C.N*) If $\Box A \in \mu \in \Omega$, and if $\mu^* \in \Omega$ is an alternative to μ , and if each free individual variable of A occurs in at least one other formula of μ^* , then $A \in \mu^*$.

Modal principles like (C.M*), (C.N) and (C.N⁺) suffice, according to Hintikka, for a minimal alethic modal logic. The notion of satisfiability of alethic modal sentences is defined in exactly the same way as the satisfiability of deontic modal sentences (see the previous section).

Hintikka (1961, 123) mentions that the alethic modal logic thus obtained is equivalent to von Wright's (1951a) system M; and that by requiring the alternativeness relation to be transitive, we obtain a stronger system that is equivalent to Lewis' system S4, and by requiring it to be symmetric, we obtain a semantical system whose syntactical twin is obtained by adding to the system M the so-called Brouwer's axiom

 $A \to \Box \Diamond A$

Further, by requiring the relation to be transitive and symmetric, we obtain a system that is equivalent to Lewis' S5. Reflexivity is not mentioned in this connection but given the commitment to the condition (C.N) it is evident that reflexivity must be included among the properties of the alternativeness relation. Hintikka (1961, 124) adds: "I shall not prove these results here". Instead he goes on to discuss the sentence

$$\exists x \Box Ax \rightarrow \Box \exists x Ax$$

which, of course, is an alethic counterpart of (49) familiar from the previous deontic discussion. Hintikka makes the same points as in the case of (49): The alethic variant is valid if the conditions (C.M*), (C.N) and (C.N⁺) are assumed, and essentially the same derivation as given in the previous section may be carried out. Again, the derivation is blocked if we replace (C.N⁺) by (C.N*) in exactly the same way as it was blocked by the replacement of (E.6) by (E.6)* in the deontic case. In other words, we can prove the validity of $\exists x \Box Ax \rightarrow \Box \exists x Ax$ if we assume that if the individual represented by a exists in μ (recall that for Hintikka a free individual variable occurring in a formula in a model set is the "formal counterpart" of an individual existing in the possible world described by the model set) then it also exists in μ^* . As we have seen, this assumption was abandoned in deontic logic and now it is abandoned in alethic logic: individuals may drop out of existence when we move from possible world to another. Hintikka contemplates the possibility of restoring the validity of $\exists x \Box Ax \rightarrow \Box \exists x Ax$ by requiring that "whatever exists in a possible state of affairs exists in all the alternative states of affairs; in short, that whatever exists exists necessarily" (Hintikka 1961, 125). He also gives a condition on model sets that "formulates exhaustively the assumption that free individual variables are transferable from a model set to its alternatives" (Hintikka 1961, 125):

(C.self=*) If a occurs in at least one sentence of μ and if μ * is an alternative to μ , then $a = a \in \mu$ *.

Two kinds of modal systems are considered in the paper. One of them, which satisfies (C.self=*), embodies the assumption that all actually existing individuals exist necessarily; the other one, which satisfies (C.N*), dispenses with this assumption. Hintikka prefers the latter: he does not subscribe to (C.self=*) and rejects the validity of

$$\exists x \Box Ax \rightarrow \Box \exists x Ax$$

just as he rejected the validity of

$$\exists x O A x \rightarrow O \exists x A x$$

in deontic logic. As we shall see in the next section, Hintikka took a different position in the context of the epistemic modalities studied in *K&B*.

Concerning the Barcan formula, Hintikka essentially repeats the remarks he made earlier in his deontic exposition: in the expanded version of the paper he considers a modification to the rule governing universal quantification, just as he did in his deontic logic

(C.U*) If $\forall x A \in \mu$ and if b occurs in the sentences of some model set which is accessible from μ , then $A(x/b) \in \mu$.

and proves the Barcan formula with the argument presented in the previous section. As before, Hintikka remarks

It is obvious, however, that the Barcan formula is unacceptable as a valid logical principle for most modalities. Clearly, what *can* exist need not always do so *actually*; (Hintikka 1969, 66)

He thereby rejects the proposed modification to the quantifier rule just as he rejected it in the deontic case. The expanded version also contains three different proofs for a theorem expressing the necessity of identity

(NI)
$$a = b \rightarrow \Box (a = b)$$

The deontic language examined in the previous section did not contain identity, but the alethic language at hand does, and no less than six conditions in addition to (C.self=*) are given concerning identity-sentences:

- (C.=) If A is an atomic sentence or an identity and if B is exactly like A except that a and b have been interchanged in one or more places and if $A \in \mu$ and if $a = b \in \mu$, then $B \in \mu$.
- (C.self \neq) Not $b \neq b \in \mu$.
- (C.self=) If b occurs in the sentences of μ , then $b = b \in \mu$.
- (C.=!) Like (C.=) except that the restriction to atomic sentences is omitted.
- (C.=*) If $a = b \in \mu$ and if μ * is an alternative to $\in \mu$ and if a and b occur in the sentences of μ *, then $a = b \in \mu$ *.

(C.N=!) If
$$a = b \in \mu$$
, then $\square(a = b) \in \mu$.

The three proofs of (NI) share the same initial steps. Suppose there is a model set μ in a model system Ω containing the counter-assumption of (NI):

1.
$$a = b$$
 $\in \mu$
2. $\diamondsuit(a \neq b)$ $\in \mu$

The latter implies, together with (C.M*)

3.
$$a \neq b$$
 $\in \mu^*$

where μ^* is an alternative to μ . Now we may derive a contradiction by applying either (C.=!), (C.=*), or (C.N=!) and show that there is no model set for the negation of (NI), and that thereby (NI) is valid. The first proof proceeds by applying (C.=!) to (2). This gives us $\diamondsuit(a \neq a) \in \mu$, which together with (C.M*) implies $(a \neq a)$ for some alternative of μ contradicting (C.self \neq). The second proof applies (C.=*) to (1) to get $a = b \in \mu^*$ which together with (3) yields a contradiction. The proof third applies (C.N=!) to (1). This gives us $\Box(a = b) \in \mu$, which together with (C.N*) implies $a = b \in \mu^*$, and again we have a contradiction.

Hintikka argues, however, that each proof is corrupted by what he considers an "illicit assumption" of unrestricted substitutivity of identicals implicitly present in the conditions (C.=!), (C.=*), and (C.N=!). In his view the three conditions are "equivalent" in the sense that the choice between them does not affect the central notions of the model set approach such as satisfiability. This is exemplified by the three proofs: all three conditions may be employed to derive the same result. Because of this "equivalence", Hintikka claims, the principle of substitutivity of identicals must be at work behind all three conditions, even though only (C.=) and (C.=!) explicitly concern substitution of terms. Further, Hintikka (1969, 68) argues that (NI) and the condition (C.N=!) are obviously invalid as "it doesn't make much sense to assume that all identities hold necessarily [...]". According to Hintikka, this, together with the equivalence claim, provides evidence against the substitutivity principle in modal contexts and against (C.=!).

Hintikka does not try to prove the equivalence of (C.=!), (C.=*), and (C.N=!), but returns to this argument in the epistemic context in $K \otimes B$. I will return to the questions of identity below.

2.7 EPISTEMIC MODAL LOGIC

In $K \not \sim B$ (1962) Hintikka investigated the satisfiability of sets of sentences involving the operators K (knowledge), B (belief), and P (epistemic possibility) in the context of model sets. The exposition does not contain any novelties in the technical sense, but the new operators require new conditions:

[...] we are led to ask how the properties of model sets are affected by the presence of the notions of knowledge and belief; how, in other words, the notion of model set can be generalized in such a way that the consistency (defensibility) of a set of statements remains tantamount to its capacity of being imbedded in a model set. What additional conditions are needed when the notions of knowledge and belief are present? (Hintikka 1962, 42)

The basic concepts are now "the agent a knows that A", symbolized by K_aA , and "it is possible, for all the agent knows, that A", symbolized by P_aA . As the quoted paragraph above indicates, Hintikka, in this new epistemic context, does not speak any longer of satisfiability and unsatisfiability but first of the *consistency* and *inconsistency* of a formula or a set of formulae, and eventually of *defensibility* and *indefensibility*, respectively. These notions are agent-relative: what is defensible for one agent may not be defensible for another. Further, instead of valid sentences Hintikka talks about *self-sustaining* sentences. Thus, to show that a set of sentences is *defensible* (satisfiable) one has to show that it is embeddable into a model system $\langle \Omega, R \rangle$. And to show that a set of sentences λ is *indefensible* (unsatisfiable), one has to show that there is no model set $\mu \in \Omega$ of a model system $\langle \Omega, R \rangle$ such that $\lambda \subseteq \mu$. Hintikka also largely suppresses the terminology of possible worlds and possible states of affairs and simply talks instead about *epistemic alternatives*, or *a-alternatives* relative to an epistemic agent *a*.

The notions $K_a A$ and $P_a A$ introduce new requirements on model systems (Hintikka 1962, 43–44). Some of them are simply counterparts or variants of their deontic or alethic precursors, for instance:

- (C.K*) If $K_a A \in \mu$ and if μ^* is an alternative to μ (with respect to the agent *a*) in some model system, then $A \in \mu^*$.
- (C. \neg K) If $\neg K_{a}A \in \mu$, then $P_{a} \neg A \in \mu$.
- (C.P*) If $P_{a}A \in \mu$, then there is at least one alternative μ^{*} to μ (with respect to a) such that $A \in \mu^{*}$.

But there are also new requirements that reflect the specific properties of knowledge and belief. For knowledge, it must be ensured that the alternativeness relation is at least reflexive and transitive with the following conditions:

- (C.K) If $K_{\mathfrak{a}}A \in \mu$, then $A \in \mu$.
- (C.KK*) If $K_a \mathcal{A} \in \mu$ and if μ^* is an alternative to μ (with respect to a) in some model system, then $K_a \mathcal{A} \in \mu^*$.

The first is needed to ensure that what the agent knows is actually the case, while the latter ensures that everything the agent knows in the state of affairs described by μ , is also known in every alternative state of affairs described by μ^* . (C.K) corresponds to the 'factivity-axiom'

$$K_{\circ}A \to A$$

while (C.KK*) corresponds to the 'knowledge-axiom' or the 'KK-thesis':

$$K_a \mathcal{A} \to K_a K_a \mathcal{A}$$

The purpose of $(C.KK^*)$ is to enforce a strong, infallible notion of knowledge based on conclusive grounds. It can be shown that in the absence of $(C.KK^*)$, there is a model set μ in a model system Ω such that both $K_aA \in \mu$ and $K_a(B \to \neg K_aA) \in \mu$, that is, in such situations, the agent knows that A but he also knows that if B is the case, he will lose the knowledge that A. Hintikka rejected this 'fallibilist' conception of knowledge and $(C.KK^*)$ ruled out model sets of this kind (Hintikka 1962, Chapter 5).

Hintikka's defense of the (C.KK*) principle makes it clear that Hintikka is concerned with *virtual knowledge*, that is, knowledge of cognitively perfect agents who are sufficiently clever to be able to carry out the implications of what they know. In accordance with this line, Hintikka's interpretation of the principles (C.K), (C.K*) and (C.KK*) is that for a cognitively ideal agent it is irrational (indefensible) to claim that the agent knows that A and, on the same occasion, to deny that A (Hintikka 1962, Chapter 5).

K&B contains many indefensibility arguments. The methodology is exactly the same as before in Hintikka (1957b) and (1961), but as already mentioned the terminology is different: Hintikka establishes that a formula A is self-sustainable (valid) by showing that $\neg A$ is indefensible (not satisfiable). Using his old familiar method, Hintikka is able to show how the epistemic counterparts of the axioms of C. I. Lewis' S4 are self-sustaining. He also gives a solution to some traditional puzzles, such as Moore's paradox. I will not assess these matters here. Instead I move on to the introduction of quantifiers in the last chapter of K&B.

The combination of epistemic notions with quantifiers and identity leads to problems analogous to those encountered in deontic and alethic quantified systems. As indicated in the introduction, these matters have been extensively debated in the literature and the exposition in $K \mathcal{C} \mathcal{B}$ raises many questions. I will not explore them in great detail here but I will return to some of them in the subsequent chapters. Let us focus here on Hintikka's notion of *knowing who* and how he perceived the difference between the logical treatment of alethic and epistemic notions.

First, Hintikka (1962, 128–129) makes explicit the existential presuppositions that underlie conditions such as (C.3) and (C.4) above. In general, these conditions are replaced by

- (C.E₀) If $\exists x A \in \mu$, then $A(x/a) \in \mu$ and $\exists x(x=a) \in \mu$ for at least one individual variable a.
- (C.U₀) If $\forall x A \in \mu$ and $\exists x (x=a) \in \mu$, then $A(x/a) \in \mu$.

I take it that the subscripts in the names of the conditions refer to the presence of zero epistemic operators. Hintikka spells out the existence of a in μ with the aid of

identity predicate. Now, Hintikka tells us, empty individual variables are disqualified as substitution instances, and we may draw an existential conclusion only from a set of sentences containing at least one existential premise. Later in the chapter Hintikka (1962, 150) calls this the "logical conservation principle".

The presence of quantifiers, identity predicate, and epistemic operators allows Hintikka to represent in his logical setting the notion *a knows who b is*. For instance, he renders "*a* knows who Mr. Hyde is" as $\exists x K_a (x=b)$. This notion introduces requirements of its own on model systems (Hintikka 1962, 160):

(C.EK=) If
$$\exists x K_a(x=b) \in \mu$$
, then $\exists x(x=b) \in \mu$.

This first condition is a quantificational version of (C.K) above that enforces the reflexivity requirement for the alternativeness relation and thus ensures the factivity of *knowing who*: if the agent knows who b is, in the state of affairs described by μ , then b exists in the state of affairs described by μ , or at least the free individual variable b occurs in the model set describing the state of affairs.

(C.EK=EK=*) If
$$\exists x K_a(x=b) \in \mu$$
 and μ^* is an epistemic alternative to μ with respect to a , then $\exists x K_a(x=b) \in \mu^*$.

This condition is a quantificational version of (C.KK*) above that enforces the transitivity requirement for the alternativeness relation and thus ensures the infallibility of *knowing who*: if the agent knows who b is, in the states of affairs described by μ , then b exists in every alternative states of affairs described by μ *. Hintikka's (1962, 159) justification of these principles is based on his decision to take *knowing who* (expressing knowledge concerning an individual) to behave logically in the same way as *knowing that* (expressing knowledge of a proposition). Thus (C.EK=) may be seen as the counterpart of the principle

$$K_*A \to A$$

for *knowing who*. And analogously, (C.EK=EK=*) may be seen as the counterpart for *knowing who* of (C.K.K*), which ensures the validity of the axiom

$$K_{a}A \rightarrow K_{a}K_{a}A$$

As pointed out earlier, this axiom ensures a strong notion of knowledge. In the same way, (C.EK=EK=*) ensures that if a knows who b is, then a will not lose this knowledge in any epistemic alternatives of a.

Hence the interaction of quantifiers and epistemic operators produces new complications, as witnessed by (C.EK=EK=*) and (C.EK=). Hintikka (1962, 161) compares these problems with their counterparts in alethic contexts, referring back to Hintikka (1961). As pointed out in the previous section constructions of the form

$$\exists x \square \dots x \dots \in \mu$$

 $\exists x \diamondsuit \dots x \dots \in \mu$

raise the question of whether an individual existing in μ also exists in the alternatives that \Box or \diamondsuit force us to consider (or at least, whether a free individual variable occurring in a model set μ also occurs in the alternatives to μ). In $K \mathcal{C}B$ Hintikka considers the analogous epistemic constructions

$$\exists x K_a \dots x \dots \in \mu$$

 $\exists x P_a \dots x \dots \in \mu$

but he interprets them in a different way. Intuitively, the quantifiers in these constructions are meant to "range" not over only individuals existing in μ but over individuals existing in μ that are also *known* (in the sense of *knowing who*, that is, identified). The problem that these constructions raise is whether an individual known by a in the model set μ is also known by a in a-alternatives to μ . And given Hintikka's notion of *knowing who* and the analogy he draws between this notion and *knowing that*, his answer is positive. That is, in the presence of epistemic operators, the following constraints on model systems are added:

- (C.E_{ep}) If $\exists x A \in \mu$ then $A(x/b) \in \mu$ and $\exists x K_a(x=b) \in \mu$ [it is assumed that A contains an occurrence of K_a or P_a and x occurs within the scope of one of them in A but not within the scope of any other epistemic operator]
- (C.U_{ep}) If $\forall x A \in \mu$ and $\exists x K_a(x=b) \in \mu$, then $A(x/b) \in \mu$ [with the same assumptions as in (C.E_{ep})].

The subscripts in the names of these conditions refer, presumably, to the presence of "existential presuppositions". Now, first, the aforementioned "logical conservation principle" also applies to conclusions in which the identity of an individual is assumed: such a conclusion may be drawn only from a set of sentences containing at least one premise in which the individual in question is identified. Second, both $\exists x K_a(...x...) \in \mu$ and $\exists x P_a(...x...) \in \mu$ fall under the incidence of (C.E_{ep}). (C.EK=EK=*) further ensures that if a knows who b is then a knows that in all a's epistemic alternatives. With the help of these principles, Hintikka shows the self-sustainability (validity) of the principle

$$\exists x K_a A x \rightarrow K_a \exists x A x$$

whose deontic and alethic counterparts

$$\exists x O A x \to O \exists x A x$$

 $\exists x \Box A x \to \Box \exists x A x$

he rejected. Here is Hintikka's argument (1962, 161). Suppose there is a model set μ in a model system Ω such that

1.	$\exists x K_a A x$	$\in \mu$	
2.	$P_a \forall x \neg Ax$	$\in \mu$	counter-assumption
3.	$\forall x \neg Ax$	$\in \mu^*$	from 2 by (C.P*)
4.	$K_{\mathbf{a}}Ab$	$\in \mu$	from 1 by (C.E $_{ep}$)
5.	$\exists x K_{\mathbf{a}}(x = b)$	$\in \mu$	from 1 by (C.E $_{ep}$)
6.	$\exists x K_{\mathbf{a}}(x = b)$	$\in \mu^*$	from 5 by (C.EK=EK=*)
7.	$\exists x(x = b)$	$\in \mu^*$	from 6 by (C.EK=)
8.	Ab	$\in \mu^*$	from 4 by (C.K*)
9.	$\neg Ab$	$\in \mu^*$	from 3 and 7 by (C.U ₀)

(8) and (9) are contradictory, which establishes the indefensibility (unsatisfiability) of the negation of

$$\exists x K_a A x \to K_a \exists x A x$$

and thereby the self-sustainability (validity) of the formula itself. Hintikka concludes the argument with the following observation:

The self-sustenance of $[\exists x K_a Ax \to K_a \exists x Ax]$ shows that there is an interesting difference between the logical behavior of the notion of knowledge and that of the notion of necessity toward quantifiers, in spite of the fact that the two are closely similar in many respects. For the notion of necessity the analogue $[\exists x \Box Ax \to \Box \exists x Ax]$ is not valid. From the fact alone that there exists an individual which cannot help having a certain property it does not follow that there necessarily is an individual with this property. For the individual first mentioned might conceivably not exist. (Hintikka 1962, 161)

Indeed, as indicated in an earlier section, the proof of the validity of the alethic variant $\exists x \Box Ax \rightarrow \Box \exists x Ax$ required the assumption that if the individual referred to by b exists in μ , then it also exists in μ^* . Hintikka rejected this assumption in the alethic and deontic logic but he now endorses it in the epistemic modal logic. This indicates that the objects that Hintikka's epistemic languages concerns are very different from the objects he takes his alethic and deontic languages to concern. I return to this issue in subsequent chapters.

Regarding identity, Hintikka presents essentially the same argument that we encountered at the end of the previous section. The names of the conditions are different and the operators occurring in the formulae are epistemic but the conclusion is the same: the principle of substitutivity of identicals is invalid in the context at hand. Assuming the substitutivity principle in epistemic contexts

amounts to assuming that "everybody knows the answers (right answers) to all questions of identity" (Hintikka 1962, 136). This is because the condition based on the principle of substitutivity (C.=!) is "equivalent" to the condition

(C.K=) If "
$$b = c$$
" $\in \mu$ then " $K_a(b = c)$ " $\in \mu$

Again, no proof is given to establish this "equivalence".

Hintikka returned to these issues in many of his writings. He deployed model sets in Hintikka (1967b), for instance, but developed entirely different kinds of conceptual tools in Hintikka (1969) and (1975), as we shall witness in Chapter 3.

2.8 COMMENTS ON THE COMPARISON

Let us return to the three explicit claims we began with: 1) Hintikka's model set semantics for quantified modal languages, introduced in 1957, was essentially the same modal semantics that Kripke first gave in 1959 and then, with significant modifications, in 1963; 2) Prior to 1959 Hintikka had proved completeness theorems for various quantified systems based on S4, S5, and M; 3) Hintikka's semantics is applicable to the philosophical problems that Kripke sought to address, namely to problems relating to semantic notions of sense, reference, intension, and extension.

I shall briefly comment on some issues regarding these claims in the light of my survey. Then I will move on to Kripke's and Hintikka's treatments of quantification.

Soundness and Completeness Proofs

To begin with the most obvious observation: no soundness or completeness proofs are to be found in Hintikka's writings 1957–1971. What may be found are statements that Hintikka calls his "soundness and completeness theorems" which, first, are not proved and, second, require clarification (see the discussion in 2.5).

Did Hintikka's Boston Manuscript or some other unpublished source contain a Kripke-style completeness proof for some system or systems, as was claimed in Copeland (2002)? This is unlikely. In Kripke's framework, we have formulae that are valid or invalid in the light of semantical model-theoretic considerations of the kind we witnessed in connection with his counter-examples to the Barcan formulae, namely considerations involving extensions of predicates, denotations of terms, ranges of quantifiers, and so forth. In addition to these, we have formulae that are syntactically provable or not provable with the chosen axioms and inference rules (or, ultimately, with the tableau construction rules), and Kripke's completeness proof is about connecting the semantically valid and syntactically provable. In Hintikka's setting, no formula may be shown to be valid by considering semantical entities in Kripke's sense (extensions of predicates, denotations of terms etc.), simply because Hintikka's framework does not

recognize such entities. Hence for a Kripke-style completeness proof there is a central piece missing in Hintikka's setting, namely a model-theoretical semantics that spells out how the semantic values of expressions contribute to the truth-values of formulae containing those expressions.

Model Sets, Tableaux, and Model-Theory

Rather than model-theoretical semantics, Hintikka's approach resembles what later became known as the method of analytic/semantic tableaux in proof theory. We may think of Hintikka's model set rules and conditions such as (C.0) – (C.6), (C.6)†, and (C.7), as well as his E-rules as rules for building up tableau-trees. These rules agree with the rules that Kripke (1959b, 4) gave for his tableau constructions. The central difference is that Kripke uses tableaux with two columns (following Beth 1955) while Hintikka's model set derivations, as witnessed above, employ only one column. The systematic integration of Hintikka's model sets into the tableau method for first-order logic was done later e.g. in Jeffrey (1967) and Smullyan (1968). We start by placing at the root of the tableau-tree the set of sets $\{\gamma\}$ (and the empty accessibility relation). Each rule is applied to a node appearing on one of the branches in the tableau and leads either to a lengthening of the current branch or to branching (via the rule (E.2), see section 2.5). A branch is closed if the set of sets of formula, with its provisional co-permissibility relation that corresponds to the branch, is such that at least one of its sets contains both a formula A and its negation, $\neg A$. The tableau is closed if each of its branches is closed. Thus what Hintikka calls his "soundness and completeness theorem" may be re-stated in the tableau jargon as

 λ is satisfiable if and only if in every tableau beginning with {λ} there is an open branch;

or, equivalently

 λ is not satisfiable if and only if the tableau construction beginning with {λ} is closed (i.e. every branch in the tableu is closed)

Recall that for Hintikka, a formula is valid if its negation is not satisfiable. Hence, for a single formula, we have

 A is valid if and only if the tableau construction beginning with ¬A is closed

Let us compare this to Kripke (1959b, 5):

THEOREM 1. *B* is semantically entailed by $A_1, ..., A_n$ if and only if the construction beginning with $A_1, ..., A_n$ in the left column and *B* in the right column is closed.

Keeping in mind that for Kripke B is semantically entailed by $A_1, ..., A_n$ if and only if $A_1 \wedge ... A_n \rightarrow B$ is universally valid and that the formula in his right column is false, then it seems, at first glance, that Hintikka and Kripke are aiming for similar results. But this is not the case. To reiterate, Kripke's notion of validity is a model-theoretical one and his theorem connects his notion of validity with the property of the tableaux. If we reconstruct Hintikka's framework in tableau terms, then it is clear that Hintikka's notion of validity and the notion of closed tableau occurring in his "theorem" are both defined in terms of tableaux. It is difficult to say what the value of such a result would be.

It is also worth noting that, in general, Hintikka's counter-examples do not have the import that Hintikka thinks they have: many of Hintikka's arguments are, in tableau jargon, tableaux that do not close (see e.g. the discussion of the Barcan formulae in the section 2.5). But what conclusions are we entitled to draw from a tableau that does not close? Let us compare once again. Kripke's (1959b, 6) proof of his "LEMMA 2" shows that from an open branch of a tableau construction beginning with A_1, \ldots, A_n in the right column and B in the left column we may extract a model $\langle \mathbf{G}, \mathbf{K} \rangle$ in which

$$A_1 \wedge \dots A_n \to B$$

is false. In other words, $\langle \mathbf{G}, \mathbf{K} \rangle$ is a counter-model to the validity of the implication. This procedure is missing in Hintikka's approach. Hintikka did not show us how to extract a model from an open tableau branch. In order to do that, we would first need a definition of the notion of a model. Kripke gave a definition, but Hintikka did not.

Two Notions of Satisfiability

Let me spell out in greater detail how Kripke's and Hintikka's key concepts of satisfiability come apart. Kripke's (1959b) semantics proper proceeds by assigning to each non-logical constant a semantic value (reference, extension) from the domain of individuals **D**. All explanatory force that this theory may be taken to possess with respect to semantic phenomena flows from this *modus operandi*, the assignment of semantic values. As witnessed, Kripke (1959b) employs these assignments to further define the notions of *valid in a model*, or better, *true in a model*, and *satisfiability*.

Hintikka's semantics, in turn, does not involve any notion comparable to assignment and hence it is not equipped to assign semantic values to linguistic items, nor does it involve anything comparable to a domain of individuals. Hintikka is not interested in "any particular interpretation", that is, in any

particular assignment of semantic values. Hence Hintikka is not in a position to define the notions of model or true in a model and thereby he cannot give truthdefinitions typical to model theory. Instead of assigning semantic values Hintikka gives syntactical conditions that sets of sentences must fulfill. If the conditions are fulfilled (or rather not violated) then, according to Hintikka, the set of sentences, or a given sentence in a particular case, is satisfiable. Hence Kripke's satisfiability and Hintikka's satisfiability are two distinct concepts. Kripke's satisfiability is a model-theoretic notion that relates an extra-linguistic entity in D and a linguistic item about that object. This relation invokes a correspondence between two sorts of entities, linguistic items and extra-linguistic entities found in D. Hintikka's satisfiability, in turn, is a relation between a sentence (or a set of sentences) and another set of sentences, that is, a model set (or a model system), which are all linguistic items. It is clear that the model set approach does not involve correspondence between two types of things in Kripke's sense as there are only linguistic things. The sentences are linguistic, the model sets and systems are linguistic, the conditions are linguistic, and these are all we are given.

Possible Worlds as Linguistic Entities

Regarding possible worlds, we witnessed interesting progress in Kripke's work: first Kripke (1959b) takes possible worlds to be complete assignments. A complete assignment is a function that maps individuals from **D** to linguistic items, but practically it is a list, and as such it is essentially bound to involve language. Hence, one might argue, possible worlds conceived in this way are, at least partly, linguistic entities. The theory as a whole is not, however, entirely linguistic as Hintikka's theory is, since it involves, for instance, the individuals in **D**. In Kripke (1963b) this picture is altered: the assignment is carried out by an external function, and thereby linguistic features of possible worlds are removed. Hence Kripke made his possible worlds and his theory less language-driven while Hintikka continued to understand semantic theory as an entirely linguistic affair.

Intensions and Extensions

Many applications of Kripke semantics are based on mapping two types of things to one another. These applications are bound to fall outside the scope of Hintikka's approach. Consider, for instance, *rigid designation*, which is, in general, a certain type of relation between a term (a linguistic item) and an individual (an extra-linguistic entity). Hintikka's semantics is incapable of dealing with such a relation: in the absence of assignments there is no way for Hintikka to fix the *designata* of terms, and in the absence of the domain of individuals there is nothing for the terms to designate (I return to this topic in a subsequent chapter). Further, one of Kripke's goals was to

[...] shed new light on questions such as the morning star paradox, and provide a semantical apparatus for sense and denotation, extension and intension, and related concepts. (Kripke 1959a, 324)

Let us bypass the question of whether or not we may construct a viable solution to "the morning star paradox" in Kripke's early framework. What is evident, however, is that Kripke's framework offers us a means to draw a distinction between extension and intension: we may, for instance, distinguish a set of individuals in **D** serving as an extension of a given predicate in a given world from its intension, a function that carries out the assignment of extensions across possible worlds (or distinguish a subset from a set of sets in **D** or a function from a another function, or how however one prefers to spell it out). For the reasons mentioned above (lack of domains and assignments) Hintikka's framework does not contain the means to define extensions of predicates or denotations of individual variables. Therefore, Hintikka cannot draw a distinction between extension and intension, and he cannot address the morning star paradox (or similar puzzles) on a lexical level by appealing to ideas such as sameness of extension and difference of intension. However, Hintikka did address such puzzles on the level of inferences (see more in section 2.9).

Philosophical Background Orientation

The general orientation of Kripke's theory is 'atomistic' or 'referential': all linguistic items are treated by the semantic theory. We may pick a predicate variable P and establish its semantically relevant features by constructing a model ϕ and by defining, say, its extension in the actual world G, by letting $\phi(P, G) = \{a, b\}$. Hintikka's orientation is 'holistic' or 'stuctural' or 'contextual': a predicate variable P does not receive any treatment as such. It does not have an extension or any other semantic value or any semantically relevant features by itself. Only in the context of a formula such as Pa may it be considered as possessing semantically relevant features, and even in this case the features are very coarse-grained: for instance, the formula Pa is satisfiable in an arbitrary model set μ unless $\neg Pa$ is present in μ .

Hintikka's semantics is a formal instrument that enables us to establish validities in the sense defined by Hintikka. Kripke's semantics is of course also a formal artifact designed, for instance, to enable proofs of meta-theoretical properties such as consistency, completeness etc., but from a broader philosophical perspective Kripke's semantics (and model-theoretic semantics in general) aims to be more than that: when we construct an intended model for a given language, model-theoretic semantics shows how the semantic values (individuals, properties, relations, and so forth) of the relevant parts of a sentence contribute to the truth-value of the sentence. This provides an insight into the relationship of language and world. Hintikka's theory is not designed to illuminate

the language-world relation in this sense. (To what extent model-theory *succeeds* to be more than a collection of set-theoretic *abstracta* is another question.)

Quantification (in Alethic and Deontic Contexts)

Hintikka's treatment of quantification in epistemic contexts is the topic of section 2.9. Here I comment briefly on Kripke's and Hintikka's treatments of quantification in alethic contexts. The most obvious difference is that Hintikka interprets quantification *substitutionally*. Clauses such as

(C.3) If $\exists x A \in \mu$, then $A(x/a) \in \mu$ for some free individual variable a.

spell out the idea that $\exists x A x$ is satisfiable if there is an appropriate substitutional instance. In this particular case, the quantified sentence is added to the model set μ whenever μ contains a substitution instance Aa for some individual variable a. Open sentences containing unbound variables suitable for quantification such as Ax do not occur and are not evaluated. Hence the notions of domain of individuals and assignment are not needed in the interpretation of quantifiers. These notions, in turn, are utilized by Kripke in his clauses for universal quantifier (iv) and (iii)* as witnessed in section 2.3 above. They display an *objectual* interpretation of quantifiers familiar from the Tarskian tradition.

The most notable feature of Hintikka's model set approach is the different rules governing quantification:

It turns out that an axiomatic and deductive system of quantified deontic logic cannot be built without limiting the applicability of the [quantifier/operator] rules [...]. (Hintikka 1957b, 23)

Neither Hintikka (1957b; 1961; 1969) nor Kripke (1963b) endorse the Barcan formulae. In Hintikka's case, their validity is blocked, first, by restricting quantifier rules, namely the rules of instantiation for the universal quantifier such as (C.4) and (E.4). We witnessed above in the discussion of deontic logic how Hintikka blocked the transition from $\forall x O \neg Ax \in \mu$ to $O \neg Aa \in \mu$. Second, Hintikka rejected rules such as (C.self=*) and replaced them with rules such as (C.N*). Rules such as (C.self=*) justify the free transferability of individual variables from a model set to its alternatives, or intuitively, an assumption that all actually existing individuals exist necessarily, while rules such as (C.N*) enforce restrictions on transferability (see the discussion in sections 2.5 and 2.6 above).

Kripke's rejection of the Barcan formulae is based on producing models that are counter-examples to them. Kripke thought that all the quantificational laws must be preserved in the context of alethic modalities, while Hintikka was convinced that the quantificational laws must be restricted (for reasons discussed earlier having to do with individuals). Kripke (1963b, 90) states that in the systems

he is concerned with, "all the laws of quantification theory" hold, including the rule of universal instantiation

$$\forall x A \rightarrow A$$

However, the both the Barcan formulae may be proved by applying the rule of necessitation to an open formula derived by this rule. Kripke (1963b, 88) refers to Prior's (1956) proof of the Barcan formula in S5 and gives the following derivation of the converse Barcan formula in the system M:

- 1. $\forall xAx \to Ay$ 2. $\Box(\forall xAx \to Ay)$ 3. $\Box(\forall xAx \to Ay) \to (\Box \forall xAx \to \Box Ay)$ 4. $\Box \forall xAx \to \Box Ay$ 5. $\forall y(\Box \forall xAx \to \Box Ay)$ 6. $\forall y(\Box \forall xAx \to \Box Ay) \to (\Box \forall xAx \to \forall y \Box Ay)$ 7. $\Box \forall xAx \to \forall y \Box Ay$
- (1) is an instance of the aforementioned axiom; (2) is obtained from (1) by the rule of necessitation for every theorem of predicate logic is necessary; (3) is an axiom of M, the "Distribution Axiom"; (4) follows from (2) and (3) by *modus ponens*; (5) is the universal generalization on (4), (6) is an instance of an axiom of quantification theory; and the conclusion, that is, the converse Barcan formula (7) follows from (5) and (6) by *modus ponens*.

If so, the systems that Kripke is concerned with are unsound: the Barcan formulae are not semantically valid but yet syntactically derivable from the "laws of quantification theory". Kripke's (1963b, 88–89) solution is to adopt the *generality interpretation* of theorems containing free variables: whenever a formula containing free variables such as (1) is asserted as an axiom/theorem, it is taken as an abbreviation of a universal generalization. This interpretation significantly restricts the operator and the quantifier rules. In particular, the initial step in the above proof, applying necessitation to an open formula derived by the rule of universal instantiation, cannot be executed. Rather, since (1) contains a free variable and must be taken as an abbreviation of

$$\forall y (\forall x Ax \rightarrow Ay)$$

If we now apply the rule of necessitation we get

$$\Box \forall y (\forall x A x \rightarrow A y)$$

To continue, we would need to derive a formula in which the modal operator occurs within the scope of the universal quantifier. But this requires a rule such as the converse Barcan formula, the very formula we are trying to prove. In general,

deriving a formula with the modal operator within the scope of a quantifier requires either an application of the rule of necessitation to an open formula or the converse Barcan formula. Kripke denies both. Prior's (1956) proof of the Barcan formula also fails, since the initial step involves the same move of applying necessitation to an open formula derived by universal instantiation:

1.
$$\forall x \square Ax \rightarrow \square Ay$$

2. $\square (\forall x \square Ax \rightarrow \square Ay)$

Necessitation may be applied only to theorems and (1) as an open formula does not qualify.

Hintikka's story of quantification has only one side: it consists exclusively of model set conditions spelling out the quantifier rules. For Kripke, his semantic treatment is one thing and the syntactic treatment is another. Both may be modified independently of each other, but with the right modifications in place Kripke is able to demonstrate that a system with classical "laws of quantification theory" under the generality interpretation is sound and complete relative to the semantics that invalidates the Barcan formulae.

(A further question that I shall not go into is whether the laws of classical quantification theory really remain intact in Kripke's interpretation. After all, his rule of universal instantiation is more complex than the classical one, and it is weaker.)

2.9 QUANTIFYING INTO EPISTEMIC CONTEXTS

Hintikka's work in modal logic went against Quine's arguments to the effect that rules such as existential generalization and substitutivity of co-designative terms are misguided in modal contexts. In particular, Quine famously argued that formulae involving quantifying into a modal context do not make sense at all. A formula such as $\exists x \Box Fx$ states, according to Quine, that an individual assigned as the value of x necessarily possesses the property F irrespectively of how x is designated. This is problematic if we keep in mind that Quine's discussion presupposed a Carnapian 'conceptual' or 'linguistic' understanding of necessity, that is, the kind of necessity exemplified by statements such as Necessarily, assassins are murderers. It is not clear how a construction such as x is a murderer under the assignment of some assassin to x can express a necessary truth in this conceptual or linguistic sense. It seems that the assassin in question should be, at least, designated with an expression that refers to or spells out the property of assassinhood in order to arrive to a necessary truth. A variable under the assignment of a value is not a type of expression that can carry out this task. Kripke (1971) solved the issue by interpreting the modal operator \square as expressing metaphysical necessity. Hence for Kripke, $\exists x \Box Fx$ states that x has a metaphysically necessary or an essential property F. This interpretation avoids the

confusion pointed out by Quine. In Hintikka's epistemic interpretation of $\exists x K_a F x$ the problem is that, in order to possess knowledge concerning an individual the epistemic subject must think or perceive x in some way or another that discriminates x from other individuals. If we state that there is an x such that a knows that x is a murderer, with x standing in for some specific individual, then "one may ask, who was this x anyway?" as Hintikka (1957a, 50) himself put it.

Hintikka acknowledges that existential generalization and substitutivity of codesignative terms do not hold uniformly in epistemic contexts. That is, one cannot always infer

1. a knows that Dr. Jekyll is a murderer (K_aM_f)

from the premises

and

- 2. a knows that Mr. Hyde is a murderer (K_aMh)
 - 3. Dr. Jekyll is Mr. Hyde (j = h)

Neither can one infer

4. There is an x such that a knows that x is a murderer $(\exists x K_a M x)$

from (2). For Quine, the failure of substitutivity in the first example indicates the referential opacity of the position occupied by the term Mr. Hyde. This feature is also responsible for the impossibility of existential generalization in the second example. Quine's solution was to restrict these rules to referentially transparent contexts. For Hintikka (1962), the failures are not failures of referentiality, that is, they are not due, as Quine sometimes seems to suggest, to the way in which our singular terms designate individuals. The source of the failures has to do rather with multiple referentiality, that is, with the fact that if a does not know that Dr. Jekyll is Mr. Hyde then a has to consider several epistemic alternatives in addition to the current one, and in some of these alternatives the proper names Dr. Jekyll and Mr. Hyde refer to two distinct individuals (Hintikka 1962, 138). For Hintikka, substitutivity of co-designative terms makes perfectly good sense in epistemic contexts, provided that a knows that Dr. Jekyll is the same man as Mr. Hyde, a requirement that Hintikka formulates as

$$K_a(j = b)$$

In an analogous way, quantifying into an epistemic context, that is, moving from

 $K_{a} \dots b \dots$

to

$$\exists x K_1 \dots x \dots$$

makes sense whenever a knows who Mr. Hyde is, that is, whenever $\exists x K_a(x = b)$ also holds (Hintikka 1962, 145). Now the difficulty alluded to above disappears because Hintikka quantifies into an epistemic context only when the epistemic subject already knows the individual in question. If we are asked in the presence of $\exists x K_a Mx$ who is this x anyway, we can always refer back to $\exists x K_a(x = b)$ and reply that x is b, that is, Mr. Hyde, an individual already known by a.

Model-theoretically speaking, Hintikka (1962) interprets sentences of the form $K_{a}(j=h)$ as expressing that the two singular terms designate the same individual in every epistemic a-alternative (and that thereby the individual in question exists in every alternative). Similarly, he interprets clauses of the form $\exists x K_a(x = h)$ as ensuring that the singular term h designates the same individual in every relevant epistemic alternative (and that the individual in question exists in every alternative) (Hintikka 1962, 152–153). However, this interpretation is ungrounded, as we shall see. Hintikka's strong conditions (C.EK=EK=*) and (C.EK=), familiar from section 2.7 above, are designed to enforce these readings when conducting inferences. In the presence of $\exists x K_a(x = b)$, the rule (C.EK=EK=*) is designed to establish $\exists x K_a(x = b)$ in all epistemic alternatives, that is, to guarantee that a knows who h is (knows to whom 'h' refers to) in all epistemic alternatives. A consecutive application of (C.EK=) establishes $\exists x(x = b)$ in all epistemic alternatives, that is, guarantees that h exists in all epistemic alternatives. With these rules, Hintikka departs from his deontic and alethic logic (and from Kripke's alethic logic) and uses them, for instance, to establish that the epistemic counterpart of the converse Barcan formula, $K_a \forall x A x \rightarrow \forall x K_a A x$, is valid in his epistemic logic: Suppose there is a model set μ in a model system Ω such that

1.	$K_{\mathbf{a}} \forall x A x$	$\in \mu$	
2.	$\exists x P_a \neg Ax$	$\in \mu$	counter-assumption
3.	$P_a \neg Ab$	$\in \mu$	from 2 by (C.E $_{ep}$)
4.	$\exists x K_{a}(x = b)$	$\in \mu$	from 2 by (C.E $_{ep}$)
5.	$\neg Ab$	$\in \mu^*$	from 3 by (C.P*)
6.	$\exists x K_{\mathbf{a}}(x = b)$	$\in \mu^*$	from 4 by (C.EK=EK=*)
7.	$\exists x(x = b)$	$\in \mu^*$	from 6 by (C.EK=)
8.	$\forall x A x$	$\in \mu^*$	from 1 by (C.K*)
9.	Ab	$\in \mu^*$	from 7 and 8 by (C.U $_0$)

Now, (5) and (9) are contradictory (the tableau closes) and the conclusion is that the converse Barcan formula is self-sustainable (valid) by *reductio*. Note that the contradiction is obtained by first deriving a substitutional instance $P_a \neg Ab \in \mu$ of (2) from which we get that b, introduced in μ , is not A in μ^* . On the other side, from (1) we know that all individuals in μ^* are A in μ^* . But since a also knows

who b is in μ^* , as (6) establishes, b exists in μ^* , by (7), and therefore by instantiating the universal quantifier with b in μ^* we get that b is A in μ^* .

As shown in section 2.3, in a Kripkean setting one cannot get a contradiction by assuming that $\Box \forall x \land Ax$ and the negation of $\forall x \Box Ax$ (i.e. $\exists x \diamondsuit \neg Ax$) are true in **G**. A contradiction is avoided because b in **G** (think of **G** as μ and of **H** as μ^*) does not have the property A in **H**, given that it does not exist in **H**. (An individual that does not exist in a world cannot have a property in that world, because the extension of a predicate is formed only from the individuals existing in that world: see section 2.3). But in Hintikka's model systems, the rules (C.EK=EK=*) and (C.EK=) have, or should have, the consequence that the individual b in a possible world described by μ also exists in a world described by μ^* and thereby falls under the incidence of the universal quantifier in μ^* . We witness here, once more, the difference between Hintikka's syntactical approach in terms of model sets and Kripke's model-theoretical approach. This time, the source of the difference lies in the principles such as (C.EK=EK=*) that Hintikka associates with his notion of *knowing who* (as witnessed in the section 2.7).

Substitution of co-designative terms, existential generalization, intelligibility of quantifying-in, and the validity of the converse Barcan formula in epistemic setting all depend, according to Hintikka, on rules such as (C.EK=EK=*) and their capability of to ensure that the singular term 'b' in $\exists x K_a(x=b)$ designates the same individual in every epistemic alternative (while also ensuring the existence of the relevant individual in all epistemic alternatives). However, it is not obvious that this is the case. In fact, none of the rules that Hintikka proposes ensures that the singular term 'b' designates the one and the same individual in every alternative possible world. This can be seen as follows.

Suppose that $\exists x K_a(x = b) \in \mu$, and that μ^* is an epistemic alternative to μ . From this we may derive, using (C.EK=EK=*) and (C.EK=), that

$$\exists x K_a(x = b) \in \mu$$

and

$$\exists x K_a(x = b) \in \mu^*$$

The most we can now get from these conditions, using the model set technique with substitutional interpretation of quantifiers, is that $(c = b) \in \mu$ and $(d = b) \in \mu^*$ for some individual variables c and d. The two conditions (C.EK=EK=*) and (C.EK=) are compatible with both the "non-rigid" interpretation of singular terms, according to which the referent of such a term may vary from world to world, and with the "rigid" interpretation, according to which the interpretation remains fixed. In other words, the non-referential semantics with its substitutional interpretation of quantifiers that the model set treatment relies on, cannot enforce that 'b' refers to one and the same individual in every relevant possible world. In our particular example, 'c' and 'b' refer to one and the same individual, say e, in

the possible world described by μ . But 'd' and 'b' may very well refer to the individual f in the world described by μ^* . However, for Hintikka's purposes, it is crucial that the singular terms come out as rigid when (C.EK=EK=*) and (C.EK=) are applied.

The problem also occurs in the context of alethic modalities: even if we have $\exists x \Box (x = b)$ in every model set, it does not guarantee that 'b' refers necessarily to the same individual in every possible world. Again, 'b' is not rigid to begin with and there are no model set rules that could guarantee rigidity. Hence the situation is similar as with the expression *The President of the United States*. The formula $\exists x \Box (x = b)$ guarantees that there is a POTUS in every possible world described by a model set but, of course, the person holding the office may vary from world to world.

Many authors have pointed out that rigid designators are not available in Hintikka's model set framework, for instance Barnes (1976), Tselishchev (1978), Sandu (2006), and Tulenheimo (2009). The first person to realize this was probably Hintikka himself. He tried to fix the issue, first by introducing new more complex model set rules in Hintikka (1969). Then he resorted to an entirely different semantic framework in the 70s, but eventually returned to model sets and tried to settle the matter once and for all by introducing rigid designators in Hintikka (1996).

In any case, Hintikka failed to provide a coherent account of quantifying-in in Kc^*B that could justify existential generalization and the substitutivity of codesignative terms in accordance with his own criteria. Further, the proof of validity of the converse Barcan formula depends on calling attention to a certain individual b found in both possible worlds described by μ and μ^* . The co-referentiality of 'b' in μ and 'b' in μ^* (which amounts to rigid designation) is supposed to enable the derivation of the contradiction. But (C.EK=EK=*) and (C.EK=) do not secure the identity in question.

I return to some of these issues in the next chapter on rigidity. Now I shall close this chapter with the following critical remarks.

2.10 INCOMPLETENESS AND UNSOUNDNESS OF HINTIKKA'S LOGIC

In the previous sections I have presented several criticisms concerning Hintikka's model set approach in modal logic. Most of them relate to the fact that Hintikka's semantic theory lacks features that may be described as model-theoretical. But is this really the case? Hintikka talks about truth, domains of individuals, extensions, and other model-theoretically loaded concepts on many occasions. True, Hintikka did not give proper semantic definitions utilizing these concepts but he did give such definitions in Hintikka (1955) for non-modal first-order logic. Is it that Hintikka actually had model-theoretical modal semantics but he simply kept it in the background without formulating it rigorously? If so, then there would be an

easy way to develop Hintikka's model set framework by enriching it with the required semantic features. Then we could recover proper models from model systems and the criticism presented above would remain rather superficial. In the following I argue that this is not the case.

Incompleteness

For a sort of case study, let us suppose that Hintikka's remarks with model-theoretical flavor in $K \dot{C}'B$ are informal fragments of a fully worked-out systematic model-theoretical semantic theory that Hintikka, for reasons unknown, left implicit. Let us call this hypothetical theory $K \dot{C}'B$ -semantics and the related hypothetical model-theoretical notions $K \dot{C}'B$ -truth, $K \dot{C}'B$ -validity, $K \dot{C}'B$ -model, and so forth. Hintikka's remarks suggest that that this hypothetical $K \dot{C}'B$ -semantics must be quite similar to Kripke's semantic theory, discussed in section 2.3 with the set of proper possible worlds, world relative domains of individuals, assignments, and the notion of true in a model and other relevant concepts defined and in place. Now, consider the notion of knowing who. According to $K \dot{C}'B$ -semantics

Existential generalization with respect to a term – say b – is admissible [...] if b refers to one and the same man in all the "possible worlds" we have to consider. If we are considering what the bearer of a knows and does not know, these "possible worlds" are the states of affairs compatible with everything he knows [...] Now b clearly refers to one and the same man in all these states of affairs if there is someone who is known by the bearer of a to be referred to by b; for then there is no state of affairs compatible with what the bearer of a knows in which b should fail to refer to him. In short, b refers to one and the same man in all the "possible worlds" we have to consider in this special case if it is true to say "a knows who b is" [...], formally $\exists x K_a(x = b)$. (Hintikka 1962, 152–153)

On the other hand, concerning substitutivity and knowledge of identity, K&B-semantics tells us that

When we are discussing what the person referred to by *a* knows we may always substitute a term (say "Dr. Jekyll") with another (say "Mr. Hyde") if and only if they refer to one and the same individual (if any) in *all* the "possible worlds" compatible with what he knows. But this is clearly tantamount to saying that the two terms are interchangeable if (and only if) the person in question *knows* that they refer to one and the same person [...] (Hintikka 1962, 139)

Now, if Hintikka's notion of *knowing who* succeeds in modeling the real relation that occurs when one person knows another person then, intuitively, it seems that

if Dr. Jekyll is Mr. Hyde and if I know who Dr. Jekyll is *and* I know who Mr. Hyde is, then I should also know that Dr. Jekyll is Mr. Hyde. If, however, it came to me as a surprise that Dr. Jekyll is Mr. Hyde then it seems that I really did not know who these people were, or rather who this person was.

Hintikka's notion of *knowing who* appears to capture these intuitions: in terms of reference, suppose that the singular terms j and h actually co-refer. Now, if I know who j is and I know who h is, then it means, according to Hintikka, that j must refer to the same individual in every epistemic alternative and h must refer to the same individual in every epistemic alternative. Since actuality is one of the epistemic alternatives and since j and h actually co-refer, they must co-refer in every epistemic alternative. If so, then according to Hintikka, j and h are interchangeable (for me), which is "tantamount to saying that [...] the person in question *knows* that they refer to one and the same person". Let us represent all this in Hintikka's symbolism:

$$[(j=h) \land \exists x K_a(x=j) \land \exists y K_a(y=h)] \rightarrow K_a(j=h)$$

Call this the *Known Identity* formula or (KI) for short. As the above considerations show, it is intuitively valid. However, as we shall see it is not provable using model set conditions. Further, if we take the idea of the existence of $K\mathcal{C}B$ -semantics seriously then, drawing from Hintikka's remarks, we may give a rigorous $K\mathcal{C}B$ -semantic argument in the meta-language for the $K\mathcal{C}B$ -validity of (KI): It may be stated in terms of comparison between what a singular term refers to in the actual world w and what it refers to in an arbitrary epistemic alternative to the actual world, w^* . Here we assume, of course, that there are proper possible worlds w and w^* described by the model sets u and u respectively.

- 1. The referent of 'j' in w = the referent of 'b' in w
- 2. The referent of 'j' in $w^* = \text{the referent of 'j' in } w$
- 3. The referent of 'h' in w^* = the referent of 'h' in w
- 4. The referent of 'j' in $w^* = \text{the referent of 'b' in } w^*$

(1) is the assumption of the truth of j = h in w. (2) and (3) are instances of rigidity assumptions that Hintikka associates with singular terms occurring in formulae of the form $\exists x K_a(x = b)$. (4) follows by the symmetry and the transitivity of the identity relation establishing the truth of j = h in w^* . Since this holds for an arbitrary epistemic alternative w^* we have the truth of j = h in all epistemic alternatives, which is by any standard semantic definition for the box-type modal operator equivalent to $K_a(j = h)$. Hintikka, without a doubt, would have agreed with such a definition. In any case, we have derived $K_a(j = h)$ on the assumptions of (j = h), $\exists x K_a(x = j)$, and $\exists y K_a(y = h)$, and thereby we have derived (KI) and established that it is $K \mathcal{C} B$ -valid.

If Hintikka's model set conditions are in line with $K \not\subset B$ -semantics, that is, if Hintikka's model set conditions state what (1) - (4) above state, we should be able to prove that (KI) is self-sustainable using the model set conditions, that is, to show that the negation of (KI) does not have a model set. But this is not the case. Suppose there is a model set μ in a model system Ω such that

1.
$$(j = b)$$
 $\in \mu$
2. $\exists x K_a(x = j)$ $\in \mu$
3. $\exists y K_a(y = b)$ $\in \mu$
4. $\neg K_a(j = b)$ $\in \mu$ counter-assumption

By equivalent transformations on (4) we get

5.
$$P_a \neg (j = h)$$
 $\in \mu$

which in turn gives us

6.
$$\neg (j = h)$$
 $\in \mu^*$ from 5 by (C.P*)

But now we already begin to see that reaching a contradiction is not possible. There is no rule to authorize transferring (j = b) from μ to μ^* and no rule to authorize transferring $\neg (j = b)$ from μ^* to μ . We could apply Hintikka's strongest conditions (C.EK=EK=*) and (C.EK=) to (2) and (3), but these only give us a series of existential formulae in μ and *knowing who* formulae with corresponding existential formulae in μ^* :

7.	$\exists x(x=j)$	$\in \mu$	from 2 by (C.EK=)
8.	$\exists y(y = b)$	$\in \mu$	from 3 by (C.EK=)
9.	$\exists x K_{a}(x = j)$	$\in \mu^*$	from 2 by (C.EK=EK=*)
10.	$\exists y K_a(y = b)$	$\in \mu^*$	from 3 by (C.EK=EK=*)
11.	$\exists x(x=j)$	$\in \mu^*$	from 9 by (C.EK=)
12.	$\exists y(y = h)$	$\in \mu^*$	from 10 by (C.EK=)

In addition to these, by (C.E_{ep}) we get

13.
$$K_a(a = j)$$
 $\in \mu$ from 2 by (C.E_{ep})
14. $K_a(b = h)$ $\in \mu$ from 3 by (C.E_{ep})

Now we can apply (C.K) and (C.K*), or (C.K.K*) at will, but we only get the identities (a = f) and (b = h) that occur in both μ and μ * without generating a contradiction. Hence there is a model set for the negation of (KI) and thus it is defensible (satisfiable), which means that (KI) is not self-sustainable in the light of syntactical model set rules but still $K\mathscr{C}B$ -valid in the light of semantical considerations.

Thereby Hintikka's model set rules and his informal $K\mathcal{O}B$ -semantics come apart. There are $K\mathcal{O}B$ -validities that are not provable with the model set rules. Hintikka's epistemic logic, that is, the collection of model set rules governing epistemic formulae, is incomplete with respect to his informally stated $K\mathcal{O}B$ -semantics. The main problem for Hintikka here is that there is no model set conditions that could state and enforce the rigidity assumptions (2) and (3) occurring in the meta-linguistic argument.

Years after Ke's B when Hintikka was distancing his work from the model sets and was moving towards world-line semantics, he introduced, in the transitional chapter "Existential Presuppositions and Uniqueness Presuppositions" of Models for Modalities (1969), the notion of modal profile together with the complex conditions (C.ind=), (C.ind=0), and (C.ind=E). Hintikka's discussion was in the context of alethic modal logic. In the chapter he discussed the necessity of identity

(NI)
$$a = b \rightarrow \Box (a = b)$$

and other related alethic formulae. Nevertheless, epistemic counterparts of these conditions would enable the proof of (KI). Here is why: Hintikka's condition (C.ind=0) states

(C.ind=₀) If
$$(a = b) \in \mu$$
,

$$\exists x[(x = a) \land \Box^{n1}(x = a) \land \Box^{n2}(x = a) \land ...] \in \mu$$
,

$$\exists x[(x = b) \land \Box^{n1}(x = b) \land \Box^{n2}(x = b) \land ...] \in \mu$$
,
then $[\Box^{n1}(a = b) \land \Box^{n2}(a = b) \land ...] \in \mu$

The idea (once again) is that (C.ind= $_0$) should guarantee the uniqueness of references of 'a' and 'b' in every relevant possible world. The superscripts indicate the different modal operators that have the variable x within their scope. If we had an epistemic counterpart of (C.ind= $_0$) at our disposal

(C.ind=
$$_{ep}$$
) If $(a = b) \in \mu$,
 $\exists x[(x = a) \land K_a^{n1}(x = a) \land K_a^{n2}(x = a) \land ...] \in \mu$,
 $\exists x[(x = b) \land K_a^{n1}(x = b) \land K_a^{n2}(x = b) \land ...] \in \mu$,
then $[K_a^{n1}(a = b) \land K_a^{n2}(a = b) \land ...] \in \mu$

we could prove (KI) by deriving, from (1), (2), and (3) above, $K_a(j = h) \in \mu$ which would contradict (4). But Hintikka did not introduce an epistemic counterpart of (C.ind=0). He contemplates the possibility but hesitates:

There are rather plausible-looking counter-examples to [(C.ind=₀)] in epistemic logic. I believe that I can nevertheless explain them away. To attempt to do it here would take us too far, however. (Hintikka 1969, 130)

From a semantical point of view, conditions such as $(C.ind=_{ep})$ are desperately needed in Hintikka's epistemic logic because without them the proof system remains incomplete. But as witnessed above, Hintikka's assumption that quantifying-in guarantees rigidity is unwarranted. $(C.ind=_0)$ and its hypothetical epistemic counterpart are also based on this unwarranted assumption. Formulae of the form $\exists x K_a(x = b)$ simply do not have the semantical import that Hintikka wants them to have. The reasons why this is so lie in the syntactical foundations of Hintikka's framework; and syntactical conditions, no matter how sophisticated, cannot fix these foundational issues. Hence there are models (in the model-theoretical sense) in which $(C.ind=_0)$ and similar principles are false, and thereby, as rules of inference, are unsound. This leads us to the next topic.

Unsoundness

As witnessed, Hintikka's "uniqueness conditions" utilizing quantifying-in such as (C.EK=EK=*) and (C.EK=) in epistemic contexts, as well as conditions such as (C.ind=0) with components of the form

$$\exists x[(x=a) \land \Box^{n1}(x=a) \land \Box^{n2}(x=a) \land \ldots] \in \mu$$

in alethic contexts, do not guarantee rigidity. From a model-theoretic point of view they do not rule out models in which the reference of 'a' varies from one possible world to another. Hence they are simply unsound and potentially authorize proofs of semantically invalid formulae.

There are also other problems in Hintikka's proof system. Before closing this chapter, let me discuss yet another semantical case that will also lead us to the next topic, namely world-lines and counterparts.

Consider Hintikka's remarks on the referential multiplicity that an ignorant epistemic subject *a* potentially confronts in a Jekyll-Hyde case.

In so far as the person referred to as *a* does not know everything there is to be known about the world he has to keep an eye on more than one "possible world," that is, he has to take into account more that one way things might actually be, as far as he knows. [...] *a* will have to keep an eye on the different ways in which names (and other individual terms) might refer to different individuals and perhaps sometimes might fail to refer altogether. If he does not know that Dr. Jekyll and Mr. Hyde are identical, he will have to heed the possibility that they might refer to different men. To put this in a slightly

different way, at least one of the "possible worlds" he will have to consider is then such that the two names refer in it to two different men. (Hintikka 1962, 139)

Linguistically there are, as Hintikka tells us, two singular terms evaluated in two possible worlds w and w^* such that in w the singular terms Dr. Jekyll and Mr. Hyde co-refer to the individual who is Jekyll-Hyde in w, while in w^* the terms do not co-refer. Hence, in model set terms, we would have $(j = h) \in \mu$ and $(j \neq h) \in \mu^*$ (with μ describing w and μ^* describing w^*). By introducing a hypothetical domain function ψ we may describe the situation in terms of individuals and domains of w and w^* as follows:

$$\psi(w) = \{b\}$$

$$\psi(w^*) = \{c, d\}$$

The singular term Dr. Jekyll refers to the individual b in w and to another individual c in w^* who is Dr. Jekyll in w^* while Mr. Hyde refers to b in w and in w^* to another individual d who is Mr. Hyde in w^* .

Now, it must be assumed that while b is the actual Jekyll-Hyde in w, somehow c is the individual who is Dr. Jekyll in w^* and also that d is the individual who is Mr. Hyde in w^* . How do we make sense of this? We need to assume that there are some kind of counterpart relations prevailing in both directions between b and c and d and d. Hence there is one individual b in w who has two counterparts, c and d, in w^* . Now the model depicts a *contingent identity*: b = b holds in w but not in w^* , in which the identity relation does not prevail between the two counterparts of b, namely c and d. As we shall witness in the next chapters, counterparts will reappear in Hintikka's work on many occasions. For now, let us consider these matters further in the context of Hintikka's alethic logic, in which Hintikka explicitly assumed contingent identity.

A number of interpretational problems emerge with such contingent identity models involving counterparts. Consider an open sentence $\lozenge Fx$. The customary semantic rules dictate that it is true of an individual b in w just in case Fx is true of b in some accessible possible world w^* . If counterparts are involved the rules must be modified to make $\lozenge Fx$ true of b if Fx is true of a counterpart of b in some w^* . Now, consider a simple model where F is a one-placed predicate only true of d in w^* . Is $\lozenge Fx$ true of b in w^* . Our modified rule has it that it is true if Fx is true of a counterpart of b in w^* . But b has two counterparts in b one that b is true of in b and one that b is not true in b has two counterparts in b one that b is true of in stipulate that b is true of b in b in b if there is at least one way to choose among the counterparts of b in b in b a counterpart that b is true of. These ideas may be explicated by assigning individualing functions to variables instead of individuals. These functions take care of the counterpart relations and pick out one counterpart of a given individual in every possible world in which the individual

has counterparts. $\lozenge Fx$ is true of b in w in this particular case if there is an individuating function of b that picks out a counterpart that is F in w^* . Now interestingly, both $\lozenge Fx$ and $\neg \lozenge Fx$ will be true of b in w. However,

$$\Diamond F_X \land \neg \Diamond F_X$$

is not true of b in w because an individuating function picks out only one counterpart per world for evaluation.

In fact, Hintikka later formulated modal semantics based on these ideas. This theory, word-line semantics, will be discussed in detail in the coming chapters. In particular, Hintikka talked about "splitting" or "diverging" world-lines that may be perhaps interpreted in many ways. But one thing we can say about them is that they emerge when an individual "splits", that is, when an individual has multiple counterparts in other possible worlds. I shall not go into to these details here. I merely mention these world-lines to indicate that Hintikka definitely invoked counterpart models later in his career.

Let us return to Hintikka's (1961) alethic logic and model sets. At this point there was no trace of individuating functions. However, if we assume that the models he tacitly considered included models with counterparts and contingent identity, then the following problem emerges: if $\forall x A$ means that A is true of everything and $\exists x A$ means that A is true of something, then $\forall x A$ is not generally equivalent to $\neg \exists x \neg A$. The above model makes this clear: b satisfies the open sentence $\neg \diamondsuit Fx$ in w when c in w^* is chosen for evaluation. Thereby $\exists x \neg \diamondsuit Fx$ (equivalent to $\neg \forall x \diamondsuit Fx$ if the standard duality relations prevail among quantifiers) is true in w. But b is the only individual in w and it satisfies also $\diamondsuit Fx$ when d in w^* is chosen for evaluation. This means that $\forall x \diamondsuit Fx$ is also true in w. Hence the duality relations between quantifiers break down. Note that individuating functions or world-lines do not automatically solve this issue. I return to this problem in the coming chapters.

Hintikka exploits the duality of quantifiers in most of his proofs. This inference rule, however, is not generally sound with respect to the hypothetical semantics we have reconstructed from informal remarks in $K \mathcal{C} B$. This, and other points made in this section, suggests that the hypothetical semantics we have tentatively attributed to Hintikka is indeed hypothetical – the semantic remarks in $K \mathcal{C} B$ are just a collection of unsystematic notes. Hintikka did not have model-theoretical semantics until *Models for Modalities* (1969). When he finally started to think his logics in model-theoretical terms, he detected many new problems. For instance, it is likely that the aforementioned rules (C.ind=), (C.ind=E), and the notion of "modal profile", as well as "straight" world-lines associated with "well-defined individuals", were introduced to re-establish the duality relations between quantifiers. They are designed to prevent the "splitting" of individuals and thereby to rule out models such as the one we have just considered.

Creation Of Epistemic Space

3 OBJECTS

This chapter studies objects of modal space by assessing the Kripkean notion of rigidity in relation to Hintikka's modal semantics. Thereby the comparative point of view established in the previous chapter will be preserved. In particular, the chapter surveys Hintikka's extraordinary struggles with rigidity and related notions from the late 50s to this millennium. I argue that Hintikka's many ambivalent remarks concerning rigidity become more comprehensible if, first, three different variants of rigidity are distinguished and, second, Hintikka's largely implicit doctrine of semantic neo-Kantianism is made explicit. The chapter is based on the previously published paper Tanninen (2019).

3.1 RIGID DESIGNATION

The characterization of rigidity given in Kripke's Naming and Necessity (1980, 48) states that an expression is a rigid designator if it designates the same object in every possible world in which the object exists and a non-rigid or an accidental designator if it does not designate the same object in every possible world. The further explications and interpretations of this notion, as well as their implications, have been widely studied and analyzed in connection with the theories of direct reference of proper names, see for instance Soames (2002), Martí (2003), and Salmon (2005).

Rigidity has also been studied outside the direct reference program. The so-called neo-descriptivists employ rigidified descriptions to capture the counterfactual variation of accidental properties of objects in modal contexts while maintaining a descriptivist theory concerning proper names, see for instance Chalmers (2004) and Jackson (1998). Here I focus on issues of rigidity that, first, are largely independent of the questions relating to proper names and, second, have not received much attention in the literature.

The notion of rigidity is essentially associated with Kripke's work but there are many reasons why we should also pay attention to Hintikka's less-known writings on the topic. Hintikka's work on the notion of rigidity was exceptional (at least) in the following respects. First, Hintikka (1957a; 1957b) was among the first to introduce a notion similar to rigidity and among the first to consider it as fundamental (for good or ill) in modal semantics. For instance, in the following passage from 1957 Hintikka describes how rigidity results from quantifying into a context of epistemic modality:

It is not quite obvious what exactly is 'meant' by (4) [There is an x such that it is known that x was the author of Shakespeare's plays]. It is certain, however, that (4) is true if there is an x who in fact was the

author of Shakespeare's plays and who, moreover, was identical with the reference of 'the author of Shakespeare's plays' in all the other possible situations compatible with our knowledge. (Hintikka 1957a, 60)³⁹

Here Hintikka describes how the variable x becomes rigid when the existential quantifier occurring in the description takes a wide scope in relation to the modal operator *it is known* (see more on *widescoping* below). Yet Hintikka kept re-evaluating the plausibility of rigidity until he presented his final critical remarks in his 2007 book *Socratic Epistemology*. Hence Hintikka struggled for five decades to come to terms with the notion.

Second, Hintikka's writings on rigidity are remarkably ambivalent. There is a *prima facie* tension between Hintikka's strict rejection of rigidity and his acceptance of something quite similar under a different name. In the mid 90s this ambivalence culminated with his publication of a long critique of rigidity and "the new theorists of reference" (Hintikka & Sandu 1995), while a few months later in another new paper (Hintikka 1996) he was happy to posit the class of rigid designators under the name "proper constants", appropriately crediting Kripke for introducing such terms.⁴⁰

Third, Hintikka struggled with the questions of rigidity in the contexts of three radically different semantic theories: with his model sets, his world-line approach, and finally with his game-theoretical semantics. These provided three different frameworks for implementing rigidity, or alternatively, for questioning its rationale.

Fourth, Hintikka's strictest criticisms seem to be directed towards a very general notion of rigidity. The issue at stake was not whether proper names, indexicals, certain descriptions, or some other expressions of natural language are rigid. Hintikka's most sceptical claim was more general, namely that rigidity has no place, or is dispensable, in the formal semantics of any applied modal language (Hintikka 1975, 28–29, also fn. 8). To understand this point, I assume below, as a working hypothesis, that it makes sense to distinguish between three different varieties of rigidity. Among other things, this means that the discussion here is somewhat abstracted from the usual debates on rigidity concerning whether this or that natural language expression is rigid or not. The central question here is not whether, say, *blue* is rigid, but whether the artifacts of our modeling that comprise our basic semantic framework should be interpreted rigidly. In addition to introducing some variants of rigidity below, I trace Hintikka's critical thoughts on rigidity roughly from the late 50s to this millennium.

In retrospect, Hintikka's negative attitude concerning rigidity came from three main sources: first, from ca. 1957 to 1969 Hintikka worked exclusively with model sets in modal semantics. As witnessed in the previous chapter, this distinctive

³⁹ Hintikka's other early remarks on rigidity in epistemic context may be found in Hintikka (1962, 150–155) and in alethic context in Hintikka (1963, 78).

⁴⁰ In Hintikka (1996) rigidity was endorsed all the way to the claim that proper names of natural language are rigid designators in the Kripkean sense.

approach provided a remarkably economic semantics for quantified modal statements by dispensing with, among other things, the notions of assignment and domain of individuals. Designation, rigid or non-rigid, was not among the concerns of this approach for obvious reasons: in the absence of assignments there was no way to fix the *designata* of terms and in the absence of domains there was nothing for the terms to designate. The viability of the approach in the inferential sense nevertheless led to an idea that Hintikka considered more or less plausible throughout his career: positing rigid designators is futile in modal semantics since whatever may be achieved by positing rigid designators as primitive may be achieved by other means.

The second source of resistance was the problem of trans-world identity, which Hintikka apparently started taking seriously after Chisholm's (1963) and Castañeda's (1964) reviews of $Ke^{i\omega}B$. Chisholm argued that Hintikka's treatment of the Fregean substitution puzzles presupposed criteria for trans-world identity while Castañeda provided similar remarks concerning the notion of self-knowledge. Hintikka's (1962, 138–141) view, as discussed in the previous chapter, was that co-designative names such as $Dr.\ Jekyll$ and $Mr.\ Hyde$ are not generally intersubstitutable in modal contexts precisely for the reason that they are not rigid: the names designate different individuals in different possible worlds unless formulae of the form $\exists x K_a(x=b)$ are present. But Chisholm argued that if this means that there are truths about $Dr.\ Jekyll$ that are not truths about $Mr.\ Hyde$ then

we presuppose that (in some very difficult sense) Dr. Jekyll in this world is identical with Dr. Jekyll in other possible worlds, and that Mr. Hyde in this world is identical with Mr. Hyde in other possible worlds. (Chisholm 1963, 793)

Chisholm continued by formulating (probably for the first time in print) his famous Adam/Noah paradox.

The questions raised by Chisholm went far beyond anything that could have been dealt with within the scope of the model set approach. In any case, Hintikka (1969) distanced himself from the model sets and eventually provided answers to Chisholm's questions by introducing his world-lines together with a novel neo-Kantian logico-semantical framework developed against a background view that in order even to perceive an individual (not to mention other epistemic attitudes) we must apply a rich arsenal of conceptual, representational, and linguistic resources such as quantification, cross-reference, and cross-identification. This led Hintikka to question the meaningfulness of any talk of individuals that overlooks cross-identification systems that keep track of individuals across the modal space. The problem with Kripke's characterization of rigidity, according to Hintikka, is precisely this: it casually mentions the *sameness of an individual* in every possible world independently of any cross-identification system and, hence, irrespective of any justification of this judgement of sameness. According to Hintikka, modal talk

of individuals makes sense only when embedded in a proper cross-identification system. The contrast is obvious to Kripke's (1980, 15–20) famous view that transworld identity is a pseudo-problem and that the sameness of individuals is just stipulated in modal talk in a straightforward and unproblematic manner.

The third source of Hintikka's rigidity criticism was his contempt for "puzzles" and his view that Kripke ultimately motivated rigidity merely by appealing to intuitions invoked by certain puzzles. Hintikka (1985) stated that it is misguided for a theorist to rely on intuitions about this or that puzzle scenario invented by the theorist herself and then generalize the data and the moral of the puzzle to serve as a basis of linguistic theorizing. One's intuitions about a given puzzle scenario are not necessarily shared by others; intuitions are subjective and simply not reliable and sharp enough. Furthermore, Hintikka (1999) claimed that contemporary analytic philosophers had acquired an intuitionistic metholodogy committed to an illusion called the "atomistic postulate" according to which "the basic input into our epistemic process consists of particular data, excluding general truths". According to Hintikka philosophical views should be systematically developed in the context of general frameworks such as model theory and modal logic, and not in the context of hitherto intuitions invoked by descriptions of some singular episodes. Hintikka (1999) claimed that Kripke's idea of rigid designation was a whim, a hasty impression based on a quick intuition invoked by a couple of referential puzzles such as his Gödel-example, and not a rigorously developed position in the contexts of model theory and modal logic.⁴¹ However, all the critical and sceptical remarks did not prevent Hintikka from utilizing rigidity whenever necessary. Recall that in $K \mathcal{C}B$ the most important topics were ultimately dealt with by appealing informally to a notion similar to rigidity. For instance, perhaps the most important epistemic notion in the book, knowing who, was modeled by appealing to an intuitive notion similar to rigidity, namely "uniqueness of reference" (Hintikka 1962, 132). I argued in the previous chapter that this notion is not, however, definable in terms of model sets.

Similarly, some kind of rigidity is clearly needed in Hintikka's world-line approach and in his game-theoretical semantics: if the cross-identification system is *physical* and the individual in question is *well-defined*, then the associated quantifiers occurring in *de re* formulae involving quantifying-in should track the very same well-defined individual across modal space (Hintikka 1969, 168–172; Hintikka 1975, 30–31; Hintikka & Sandu 1995, 249). A straightforward way to understand this in terms of the Kripke semantics familiar from the previous chapter is that the assumed models come with constant or at least with overlapping domains, that is, the domains of different points of evaluation (possible worlds) consist (at least partly) of the very same objects and that the assumed variable assignment is a constant function that picks out the same object in every possible world. But if so, it seems that the semantic value of the variable in question is just this same well-defined object in every possible world, and hence

⁴¹ This claim was, of course, highly provocative.

the variable must be rigid since it indeed picks out the very same object in every point of evaluation. Hence there appears to be a contradiction in Hintikka's approach: some kind of rigidity was desperately needed for many purposes, and yet Hintikka declared categorically that it should be "put out of its misery" (Hintikka & Sandu 1995, 281).

A common way to come to terms with some of these discrepancies in Hintikka's writings has been to claim that rigidity may be "defined", "accounted for", or "analyzed" in Hintikka's framework without positing actual rigid designators. These claims have been put forward in varying forms and strengths, for instance in K&B, Hintikka & Sandu (1995), Perry (2009), Tulenheimo (2009), and Pietarinen (2010). After reviewing the relevant material, I claim below that this is a misguided idea. I argue that there are semantic frameworks that do not accommodate rigidity, and to argue that such frameworks are the way to proceed in applied modal semantics may lead to viable criticism of rigidity. But if one operates with a framework that is sufficiently strong to define rigidity, then the crucial moves have already been made: in a sufficiently strong semantic framework for implementing rigidity it has been assumed, first, that the objects that one's language is about are such that it makes sense to consider them as being the same from world to world and, second, it has been assumed that one's semantics contains the means to track those objects from one world to another. To assume these is to assume rigidity in one specific sense. One may employ this specific rigidity to describe other, perhaps more general, forms of rigidity but the profound criticism that Hintikka seemed to have in mind towards rigidity becomes questionable after accepting the assumptions above.

3.2 VARIETIES OF RIGIDITY

Let me first sharpen the intuitive conception of rigidity implicit in Kripke's (1980) characterization of the notion. If we operate with the semantics familiar from the previous chapter, that is, with models of the form

$$\langle \mathbf{K}, R, \mathbf{D}, \phi \rangle$$

where **K** is a non-empty set of points of evaluation or possible worlds, R is an accessibility relation on that set, **D** is a non-empty set of individuals, and ϕ is an interpretation function, then a term t is rigid just in case for every possible world $w \in \mathbf{K}$ and for every $v \in \mathbf{K}$ such that wRv, the interpretation function ϕ picks out in the case of t the same individual d in **D** at w and v. This chapter is mainly concerned with the case of the simplest rigid terms, namely variables under assignments of values. Variables are not expressions of any natural language but expressions of a formal language, and they are the basic artefacts of our modeling. One way to make a variable rigid, standardly employed in giving an objectual treatment of quantifiers, is to posit a constant interpretation function, or more

precisely, a constant variable assignment function, that has the same value in every point of evaluation, that is, a function that picks out the same object in every possible world. Kripke's (1959b) interpretation of a free variable that assigns to every free variable of the formula \mathcal{A} an individual from \mathbf{D} is precisely this kind of rigid interpretation, as was emphasized in section 2.3. The interpretation in Kripke (1963b), where a variable is assigned an individual from the union of the domains of all possible worlds, is also a rigid interpretation (see the relevant clauses given in section 2.3).

Rigidity, as defined, has many uses in theorizing. To appreciate Hintikka's critical points on rigidity, I assume first that it makes sense to distinguish between three further variants of rigidity, or better, to distinguish between three different contexts in which rigidity in the sense above may be entertained (with each context creating different commitments). In absence of better terms, I call them *formal rigidity, semantical rigidity*, and *linguistic rigidity*.

Kripke (1980, 3-4) referred to the first when he stated that rigidity is "a possibility that certainly exists in a formal modal language". Kripke's thought here seems to be that a mathematical logician, for instance, may stipulate rigid designators at will in order to study how assumptions concerning available designators affect, say, the meta-logical properties of some systems. The minimal requirements for a term to be formally rigid are, first, that the term is a term of a formal language and, second, that it is stipulated by a logician to be rigid in the sense above. Entertaining this type of rigidity does not commit one to any claims outside mathematical logic and in particular it does not commit one to any views concerning natural language whereas entertaining the last type of rigidity linguistic rigidity - does. It is the most frequently discussed variant of rigidity, having to do with specific questions in natural language semantics such as whether natural kind terms are rigid and, if they are, what exactly they rigidly designate.⁴² Hence, to entertain linguistic rigidity is to claim that this or that natural language expression is rigid. For a term to be linguistically rigid, the idea is that, first, it is not a mere artefact such as a variable but an actual expression of some natural language and, second, it is rigid roughly in the above sense by virtue of the semantic conventions prevailing in a linguistic community. Of course, now the set of worlds, the domain, and the interpretation function are not mere set-theoretical abstracta but something that **K**, **D**, and ϕ are taken to be abstractions of, such as alternative courses of history, people who are discussed, and prevailing semantic conventions. Kripke's view concerning a proper name such as Gödel is that, as an expression of a natural language, it is governed by a semantic convention according to which a proper name keeps designating the same person when discussing possible alternative histories of the person in question. The mere possibility of formal rigidity alone does not justify such a conclusion concerning natural language (see Kripke 1980, 4). A standard practice, for instance in the

⁴² This topic was debated extensively after Soames' 2002 book.

research program of translational semantics, is to employ formal rigidity to model linguistic rigidity by appealing to "hermeneutic principles" (as Sosa 1996 has called them) such as: If a proper name in ordinary natural language is rigid then it may correctly be represented logically by an individual constant governed by a constant interpretation function (Sosa 1996, 38).

To entertain the intermediate notion, semantical rigidity, is to accept formal rigidity and to accept its applicability to natural language semantics in some contexts but to deny linguistic rigidity in the above sense, that is, to deny the rigidity in the case of natural language terms, and in particular abandon the hermeneutic principles above that assimilate some natural language terms to rigid terms of a formal language. However, formally rigid terms, such as variables, are needed to model how some propositions and some inferences display a rigidity phenomenon resulting from an interplay of certain terms, concepts, and logical forms, since these must be recognized in applications of philosophical logic to reasoning with alethic modalities, epistemic attitudes, intentional states, deontic concepts, etc.

Is this trichotomy sustainable? That remains to be seen, but it is useful for the purposes at hand. The current objective is to understand Hintikka's thinking on rigidity and my claim is that it presupposes something like the trichotomy above. The correctness of the distinction is a separate question. Anyhow, to clarify the distinction, consider the following caricatures: a direct referentialist, a widescoper desciptivist, and a counterpart theorist. If the first held the popular view among the direct referentialists that all directly referential terms in natural language are rigid, then she would entertain linguistic rigidity and go all the way with rigidity, that is, she would accept all varieties of rigidity. Then consider a typical widescoper descriptivist. 43 If she held the standard descriptivist view that all singular and general terms in natural language are synonymous with descriptions, then she would not entertain linguistic rigidity (since no description as such is rigid). But she would entertain semantical rigidity in explaining by widescoping how some terms (which are not rigid per se) interact with alethic modalities in reasoning and in some propositions. Hence, according to her, no natural language term is rigid as such, but some inferences and propositions display a semantical rigidity phenomenon when terms interact with modal adverbs. For example, a 'Kripkean reading' of the statement

Someone who won might not have won

may be recovered by widescoping, that is, by explicitly considering the description someone who won as taking a wide scope in relation to the modal expression might have. With the tools of quantified modal logic this reading may be symbolized as

$$\exists x (Wx \land \Diamond \neg Wx)$$

⁴³ Such as the one introduced in the second section of Caplan (2005).

and according to standard semantics it is true in the world w if and only if there is a winner i in w such that there is an accessible possible world w^* such that i did not win in w^* , that is, just in case the very same individual i who won in w is to be found as a loser in w^* . Anyone using the sentence above must mean this since the other option would be, as Dummett (2000, 114) put it, "such a stupid thing to say": the narrow scope reading in which the description takes a narrow scope relative to the modal adverb, namely

$$\Diamond \exists x (Wx \land \neg Wx)$$

is always false since its truth would require there being an accessible world with a winning loser, an apparent contradiction in terms. Hence, for a widescoper the recognition of the Kripkean reading of the sentence above involves entertaining semantical rigidity but not linguistic rigidity since none of the expressions comprising the sentence above are rigid as such. A rigidity phenomenon nevertheless must be recognized at the level of variables in the logical representation of such statements in order to avoid stupidity when the description, the modal adverb, and the logical form in question interact the way they do above.

A counterpart theorist, in turn, would not accept the semantic story of the widescoper. No winner in w is identical to any loser in w^* (or to any other individual in w^* for that matter). According to a counterpart theorist the domains of the possible worlds are totally disjoint. A winner in w and a loser in w^* may, at best, be counterparts. Hence linguistic, semantical, and formal rigidity are all ruled out because there are no suitable individuals to be designated rigidly. A counterpart theorist would probably not object if one, for mathematical purposes, stipulated models with constant domains and entertained formal rigidity. But the resulting semantics would not be counterpart semantics.

The strictest view attributable to Hintikka refrains from entertaining rigidity in any of these contexts. As shown above, the model set approach dispensed with assignments and domains of individuals. Hence Kripke's statement that rigidity is "a possibility that certainly exists in a formal modal language" is not true in this particular case (namely, in the case of the formal modal language governed by the model set semantics). The possibility of rigidity requires a suitable semantic framework that posits objects and relations between linguistic items and those objects. It is not possible in the formal language of *K&B* to entertain even formal rigidity, since model sets do not provide any *designata* that could be designated rigidly or non-rigidly. The most liberal view attributable to Hintikka, in turn, accepts all three variants of rigidity, since he accepted linguistic rigidity in his notorious 1996 paper. Below, I argue that in order to understand Hintikka's struggles with rigidity it is instructive to suppose a notion similar to semantical rigidity above. Hintikka went back and forth with rigidity, but at the end of the day it was an intermediate view between formal and linguistic rigidities that Hintikka

wanted and needed to establish. By and large, the key question for Hintikka was whether semantical rigidity should be accepted or not.

3.3 RIGIDITY IN MODEL SETS

As witnessed in the previous chapter, Hintikka's model set technique dispensed with the notions of assignment, domain of individuals, truth, and model. The theory does not posit objects other than linguistic entities, and hence there is virtually nothing to be said about the individuals that Hintikka's modal languages concern.

Many authors, including Hintikka (1996) himself, have pointed out that there are no rigid designators in Hintikka's early model set framework, see for instance Barnes (1976), Tselishchev (1978), Sandu (2006), and Tulenheimo (2009). This is a correct observation as long as it is strengthened to cover two further interrelated aspects: first, not only rigid designators but also designators of any kind are non-existent and, second, for that reason methods such as widescoping cannot be used to recover rigidity.

The candidates for designators in Hintikka's model set framework were the "bound variables" (x, y, ...) and the "free individual variables" or "free individual symbols" (a, b, ...), as they were called in $K \mathcal{C} B$. It is important to note that "free individual symbols", despite their appearance, are not individual constants. Individual constants would designate (or be proper names of) particular individuals. Since Hintikka was not, in general, "interested in truth under some particular interpretation of logical formulae", he did not need individual constants to begin with, and hence the terms of his language, namely the both types of variables, were not recognized as designating anything in particular. Hintikka claims, however, that both types of variables are "ranging over" individuals, but since the apparatus of model sets does not assign any semantic values such as referents to these terms, they cannot be designators. In order for a term to be a rigid designator, it must first be a designator. Hence terms in Hintikka's modal languages are neither designators nor rigid designators as such. Hintikka acknowledges the latter fact since he takes both types of terms as fundamentally "multiply referential", that is, as non-rigid. Hence Hintikka rules out linguistic rigidity, as defined above, right at the outset.

As witnessed in the previous chapter, Hintikka wanted nevertheless to establish semantical rigidity by invoking formulae such as $\exists x K_a(x=b)$, accompanied with rules designed to secure the rigidity of the free individual symbol. Hintikka (1962, 156) sought to distinguish between formulae that display modal "referential multiplicity" such as K_aFb (where b is merely an "individual variable" meant to designate random individuals that have the property F) and formulae that display "uniqueness of reference" such as $\exists x K_a(x=b)$, that is, formulae in which semantical rigidity in the aforementioned sense is imposed by appealing to the

assumed rigidity of the variable. However, nothing in the model set apparatus supports this distinction.

Hintikka's strategy here is clearly widescoping, but in this case it has no impact. Widescoping works because it posits a constant variable assignment function that picks out, for the value of x, the same individual in every possible world. With his substitutional interpretation of quantifiers and his notion of coherence-based satisfaction (see the discussion in section 2.8) Hintikka cannot appeal to such functions, and thereby semantical rigidity remains out of reach, even via widescoping. In his aforementioned 1996 paper, Hintikka temporarily abandoned his rigidity criticisms and posited rigid designators under the name proper constants. If such a rigid constant b were available then $\exists x K_a(x = b)$ would enforce uniqueness of reference because the variable would inherit rigidity from the rigid constant. This would reverse the explanation of uniqueness, since Hintikka's original idea was that the free individual symbol b inherits rigidity from the variable x via quantifying in. However, with proper constants in the picture a sentence such as K_aFb alone would imply that a knows who b is. One of Hintikka's guiding ideas in $K \mathcal{O} B$ was that this is not the case: $K_a F b$ must always be supplemented with $\exists x K_a(x = b)$ as an extra premise in order to conclude that a knows who b is. Positing rigid proper constants would alter Hintikka's epistemic logic in fundamental respects, and Hintikka probably re-established his critical position for this reason.

To conclude, the model set approach, as it stands, does not support linguistic, semantic, or formal rigidity. According to the traditional understanding substantial matters of reference and truth cannot be extracted from a syntactic proof-system alone. I argued here that these matters cannot be extracted from model sets alone.

3.4 WORLD-LINES AND NEO-KANTIANISM

Hintikka abandoned model sets, at least temporarily, and acquired model-theoretical semantics in two central chapters of *Models for Modalities* (1969). The reasons for this change were the aforementioned questions of cross-identification raised by Chisholm and others, and Hintikka's general discontent with his earlier linguistic approach. Hintikka eventually spelled out his dissatisfaction in the chapter "Carnap's Heritage in Logical Semantics" of *Intentions of Intentionality* (1975), which despite its title, was more a critical assessment of Hintikka's own model set approach than a study of Carnap's philosophical legacy. According to Hintikka, Carnap's (read Hintikka's) semantics was not possible worlds semantics at all due to its linguistic nature and its syntactic-axiomatic orientation.

In any case, instead of the earlier linguistic surrogates, Hintikka (1969, 87–90) now pursued by and large the same ideas (namely, the ideas presented in $K \circ B$) by invoking notions such as "satisfaction" in the Tarskian correspondence-based sense and "truth in a model". Hintikka starts by laying out a semantics technically similar to Kripke (1963b) with the set of possible worlds Ω , an alternativeness

relation on that set, domain of individuals **I** (from which world-relative domains $\mathbf{I}(\mu)$ may be recovered), and interpretation function ϕ . Hintikka gives a truth definition in the form of a semantic clause typical to model-theory for formulae containing a modal belief operator B_a :

 $B_{a}p$ is true in a possible world μ if and only if p is true in all the alternatives to μ . (Hintikka 1969, 94)

Note that μ and its alternatives here are not model sets but members of Ω , that is, extra-linguistic non-representational possible worlds. Regarding rigidity, Hintikka (1969, 96) restated his old view that singular terms are not rigid but rather multiply referential. Hintikka's general position at this point (1970, 410) was that indexicals, demonstratives and proper names of natural language are not rigid. Thereby linguistic rigidity was ruled out right at the outset.

Then Hintikka moves on to address quantification in propositional attitude contexts within his model-theoretical framework. At this point Hintikka departs from Kripke's approach. According to Hintikka (1969, 96-97), a Kripke-style treatment of quantification (see section 2.3) is untenable in epistemic contexts due to problems emphasized by Quine, namely problems relating to existential generalization and the substitutivity of co-designative terms (see section 2.9). Hintikka restates his earlier conviction that these problems may eventually be solved by requiring that the terms involved refer to the same individuals in every accessible possible world. This requirement is unsurprisingly expressed in the case of belief by the formulae of the form $\exists x B_a(x = b)$. Hence Hintikka aims to establish semantical rigidity by wide-scoping just as before. But what does it mean for a term to refer to the same individual in two or more possible worlds? As witnessed, the model set technique does not provide an insight into this question. On the other side, Kripke's objectual interpretation of quantifiers would provide a straightforward account of objects and identity, but it would also licence, for instance, existential generalization without restrictions in epistemic contexts. Hence Hintikka sought to deal with quantifications by making sense of trans-world identity in epistemic contexts. He drew inspiration, once again, from Carnap's (1947) work, and his notion of individual concept:

The way to do so [i.e. to make sense of trans-world-identity] is to postulate a method of making cross-identifications. One possible way to do so is to postulate a set of functions \mathbf{F} each member f of which picks out at most one individual $f(\mu)$ from the domain of individuals $\mathbf{I}(\mu)$ of each given [...] μ . We must allow that there is no such value for some [...] μ . In other words, $f \in \mathbf{F}$ may be a partial function. Furthermore, we must often require that, given $f_1, f_2 \in \mathbf{F}$, if $f_1(\mu) = f_2(\mu)$ then $f_1(\lambda) = f_2(\lambda)$ for all alternatives λ to μ . In other words, an individual cannot 'split' when we move from a world to its alternatives. (Hintikka 1969, 100)

Hintikka (1969, 168–169) stressed that all modal talk presupposes a "method of individuation" or a system thereof:

A quantifier that binds (from the outside) a variable occurring in a modal context does not make any sense without such a method of individuation, and its meaning is relative to this method. (Hintikka 1969, 169)

Hence the idea seems to be that the functions in $\bf F$ are methods of individuation, and these methods are prerequisites of quantification in modal contexts. Therefore the functions in $\bf F$ must first be established before we may evaluate quantified formulae. The functions are partial, that is, they may or may not have a value in a given world and, second, not every function from possible worlds to individuals is a member of $\bf F$. These functions are supposed to cross-identify, that is, pick out "manifestations" of the one and the same individual from one world to another. A function that chooses values arbitrarily from world to world, say Rudolf Carnap at μ and the number 2 at μ^* , does not individuate any individual, since assumedly there is no such individual that could manifest itself as R. Carnap in one world and as the number 2 in another world.

The functions in F were precursors of world-lines, which were properly introduced in the subsequent chapter "On the Logic of Perception" of the same volume. The most original and interesting aspect in Hintikka's new approach was that he attempted to establish semantics for four different quantifiers by appealing to different kinds of world-lines (the perspectival world-lines in the case of perspectival quantifiers Ex, Ax and the physical or public world-lines in the case of physical or public quantifiers $\exists x, \forall x$). However, Hintikka's remarks concerning the nature of these world-lines and the exact meaning of his quantifers remained remarkably vague: for instance, he never gave model-theoretical truth definitions for his four types of quantified formulae. Instead, Hintikka provided many informal imaginative explanations. The above notion of "splitting" was one guiding idea in drawing a distinction between the two types of world-lines. Another one was "diverging": in the physical/public quantification it seems that we must require that the world-line does not split or diverge, while in the perspectival quantification these seem to be allowed. It is rather difficult to say what these notions mean exactly, and hence it is also difficult to say what really is the difference between the two types world-lines and the associated quantifiers. On many occasions Hintikka states that in physical (public) quantification the associated world-line picks out the same object in all possible worlds, while in perspectival quantification the world-line picks out different objects depending on perspective. But this is uninformative since we were told that the apparatus of world-lines should explain these very notions of sameness and difference in modal contexts. If we say that the physical quantifier differs from the perspectival quantifier in that the former picks out the same object in every possible world while the latter does not, then we assume that the

question of sameness has been already settled. But Hintikka's main point here is that the question of sameness requires a proper explanatory answer. The general informal picture put forth, for instance in Hintikka (1970), was something like this:

There is a distinction between how things are (or at least how the majority of epistemic agents take them to be) and how they may appear to a particular epistemic subject. Physical or public well-established world-lines emerge when successful communication and correct use of expressions generate coherent epistemic attitudes towards specific well-defined individuals individuated via a physical or public mode of individuation in objective ontology that epistemic subjects can agree on. A model depicting such an epistemic state contains, in Hintikka's metaphor, a straight world-line, indicating that a speaker associates the same correct referent with a given expression in all possible worlds compatible with what the speaker believes. With physical or public world-lines and quantifiers, it would be appropriate to talk about "uniqueness of reference". Perspectival, perhaps poorly established, world-lines, in turn, may imply indefinite or deviant use of expressions, vague or incoherent epistemic attitudes, and obscure subjective ontologies with individuals defined via a perspectival mode of individuation. A model depicting such an epistemic state contains bending or diverging or splitting world-lines that indicate, for instance, that even though a speaker thinks that her references, individuals, and attitudes are perfectly in order, the order prevails merely from her perspective. However, from a more objective (public), perhaps physical point of view we might observe that the speaker is not fully informed or is confused, that she has not identified individuals proper but mere roles played by the individuals in her epistemic attitudes and any individuals that appear similar to her can play those roles for her. Hence the world-lines bend and pick out different individuals in different possible worlds. These individuals may seem identical from the subject's perspective, but 'in reality', or from a public or physical viewpoint, they are in fact different individuals. With perspectival quantifiers and world-lines it would be appropriate to talk about "referential multiplicity". Hence at least two notions of sameness emerge here: to say that there is a "straight" world-line associated with physical or public quantification is to say that in one sense the same individual is picked out in every relevant possible world, while to say that there is a "bending" or "diverging" or "splitting" world-line is to say that the same individual from one perspective (of an epistemic subject) is picked out in every possible world.

From a more practical point of view, Hintikka probably realized that in $K \not\subset B$ he had conflated two notions that he should have kept distinct: *knowing who* and *knowing of someone that she is F*. The latter does not imply the first. In $K \not\subset B$ one can only possess knowledge of an individual if one knows who that individual is. That does not always seem to be the case. Consider the following scenario from a TV-series: a passerby, let's say a, happens to witness a murder and gets a look at the murderer. Later a is called to the police station to identify the murderer in a lineup. Now a may pick out the murderer from the line without having any idea who the murderer is. Of course, a could say that he is the murderer but that would be just

to restate that a knows of the man that he is the murderer, not that a knows who he is. Hence it is evident that a knows many things about the murderer without knowing who he is. In $K \mathcal{O} B$ the only way to describe a's epistemic state directed to an object, the murderer, would be $\exists x K_a M x$, but that implies the presence of $\exists x K_a(x = b)$ for some b which, in turn, for Hintikka means that a knows who x is. $K_a \exists x M x$, in turn, would not express a's knowledge of the murderer but knowledge of the general proposition that someone is a murderer. Now with Hintikka's new tools we may utilize the perspectival quantifier and describe a's epistemic state as ExK_aMx which does not imply that a knows who x is, that is, $\exists x K_a(x = b)$. Perhaps a weaker statement is implied, namely $ExK_a(x = b)$. In any case, the perspectival world-line associated with Ex (at least according to one interpretation) does not invoke "uniqueness of reference". The value of the function is allowed to vary from one possible world to another. In other words, a has individuated the murderer merely by visual impression and any person with sufficiently similar looks could be picked out by the perspectival world-line in the alternative worlds.

Now the question is how to make all that precise at the level of a semantic theory. In the absence of Hintikka's explications Niiniluoto (1979; 1982) provided roughly the following model-theoretic account of quantified formulae. In Niiniluoto's symbolism O_a is a modal attitude operator such as knowledge or belief relative to a subject a, while $\mathcal{O}_a(\mu)$ is the set of \mathcal{O}_a -alternatives to the actual world μ in Ω that a subject relative accessibility relation R_a provides access to.

Physical/Public Quantification

 $\exists x O_a Fx$ is true in μ if and only if there is a physical/public world-line f on $\mathcal{O}_a(\mu)$ such that, for all $\mu^* \in \mathcal{O}_a(\mu)$, $f(\mu^*)$ satisfies Fx in μ^* .

Perspectival Quantification

 ExO_aFx is true in μ if and only if there is a perspectival world-line f on $\mathcal{O}_a(\mu)$ relative to μ and to a such that, for all $\mu^* \in \mathcal{O}_a(\mu)$, $f(\mu^*)$ satisfies Fx in μ^* .

As we see, Niiniluoto takes both types of quantifiers as ranging over world-lines, in other words, he interprets quantifiers intensionally. Now we may evaluate quantified formulae given that a system of world-lines has been established beforehand. Two problems emerge: first, Niiniluoto's definitions characterize the difference between the two types of quantifications by appealing to different types of world-lines. But what exactly is the difference between a physical/public world-line and a perspectival world-line? We do not know enough of world-lines and the associated criteria, and consequently we do not know how the two types of quantified formulae differ in meaning. Second, several of Hintikka's remarks go against the idea of intensional interpretation of variables (albeit some his remarks evidently support this idea, see below).

Perry (2009) takes the referential remarks in Hintikka's writings such as "the values of the variables of a formal language must be real, full-fledged individuals" (Hintikka 1967a, 38) as suggesting the extensional interpretation of quantifiers: in an existentially quantified formula specifying a propositional attitude concerning, say Dr. Salazar, the value of x must be Dr. Salazar. If an individual such as Dr. Salazar is assigned as the value of the variable in any formula displaying quantifying-in such as $\exists x O_a Fx$ then the truth of such a formula in a standard Kripke model requires there to be a constant variable assignment function that tracks Dr. Salazar in every possible world introduced by the modal operator. This interpretation would fulfil Hintikka's demands in the case of physical/public quantification: Hintikka (1969, 103-104; 1970, 31) held onto his old idea that $\exists x K_a(x = b)$, now understood as an instance of physical/public quantification, enforces semantical rigidity by ensuring that b refers to the same individual in every possible world. But this interpretation introduces problems in the case of perspectival quantifiers: the distinctive formal outcome of perspectival quantification is supposed to be that the value of the variable may vary from world to world (Hintikka 1974, 218-219). If all quantification must be first-order objectual quantification, then the individual assigned as the value of x bound by a perspectival quantifier must be quite peculiar, since its identity would vary from one world to another, as Perry (2009, 380) acknowledges. At least, such an individual is hardly "real" and "full-fledged".

Another option would be to first interpret physical quantification extensionally, as Perry suggests, and then to interpret perspectival quantification as intensional and end up with something like what was introduced in Holliday & Perry (2014), in which standard extensional quantifiers range over objects and intensional quantifiers range over functions that represent *epistemic roles*. But this would, in the case of physical/public quantification, undermine Hintikka's view concerning the importance of a cross-identification method:

Each possible world contains a number of individuals [...] with certain properties and relations to each other. We have to use these properties and relations to decide which member (if any) of a given possible world is identical with a given member of another possible world. Individuals do not carry their names in their foreheads; they do not identify themselves. We cannot – even counterfactually – observe bare particulars, only particulars clothed in their respective properties and interrelations. (Hintikka 1970, 28)

According to Hintikka, the meaning of a quantified formula, including physically/publicly quantified formulae, is relative to an individuation method that distinguishes the attitudinal target from other individuals by appealing to the target's properties. If we quantify objectually into an epistemic context $\exists x K_a(x = b)$ with respect to a standard Kripke model, then neither properties nor relations considered, nor any identification methods employed by the epistemic subject, are

relevant to the truth or the falsity of $\exists x K_a(x = b)$ in the sense that Hintikka seems to demand above. The only relevant thing is whether b itself, regardless of its observable properties, is assigned as the value of x in every relevant possible world. If so, the meaning of $\exists x K_a(x = b)$ is not relative to any method of individuation.

Hence there are a number of problems: how do we interpret the quantifiers in relation to methods of individuation? What is *sameness* for Hintikka, and how does it relate to rigidity? How do we make sense of Hintikka's critique of rigidity?

At this point it may be instructive to assess Hintikka's neo-Kantianism, a theme almost completely neglected in the literature. "Semantic neo-Kantianism" is how Hintikka (1969) described his position, at least "as far as our thinking is concerned". What does semantic neo-Kantianism mean in the case at hand?

My suggestion is that it means that Hintikka was, or he should have been, an *instrumentalist* concerning the relational models of model-theoretical semantics. In order to clarify his world-line approach and his rigidity critique, and in order to provide a proper alternative to the Kripkean understanding of variables, Hintikka should have stated clearly that he had resigned from the Kripkean *realism* concerning the relational models in modal semantics.⁴⁴

In the context of (naive) Set Theory it is unproblematic (and necessary) to speak about individuals and say, for instance, that the element a is a member of the sets A and B. The semantics of modal languages is defined in terms of Set Theory and in order to give truth-conditions to modal formulae we use the simplistic parlance of Set Theory and state, for instance, things that imply that one and the same object a has, for instance, the property P in the possible world w and in the possible world v. In the Kripkean realistic orientation, modal semantics inherits this simplicity from Set Theory since model structures or frames may be superimposed on reality: or better, reality is the frame when relational semantics is applied to natural language sentences containing modal notions in the sense that the individuals of which the domains are composed are the very objects surrounding us.

According to an alternative view, namely the view that Hintikka should have formulated more clearly, this simplistic picture would be just an idealization. We say things that imply that a has P in w and v but it does not mean that a is really the same object in w and v. The object a in w and the object a in v are, at best, counterparts. This is where world-lines come into the picture. The counterpart relations consist of world-lines and Hintikka suggests that there are two main kinds of counterpart relations. Intuitively, two objects may be counterparts in several respects. Two objects may be physical counterparts and thus they resemble one another closely in all physical respects. This similarity may be objective or at least inter-subjective (public) in the sense that many agents agree on that. But two objects may also be epistemic counterparts relative to a subject's experience in the

⁴⁴ See Stalnaker (2016) for the distinction between the instrumentalists and the realists in the case of applied modalities.

sense that they are the same as far as the subject knows. These perspectival counterparts appear similar or even the same from a certain point of view. For instance, two objects may be very different, yet indistinguishable when perceived from a certain distance. In such case the objects are perspectival and *perceptual* counterparts. Or some objects may be perspectival and *mnemonic* counterparts, as Hintikka explains:

[Perspectival method] relies on the role of the person whose attitudes we are discussing. Let us suppose that that individual is myself and that the propositional attitude in question is memory. Then my own firsthand memories of persons, times, places, and objects create a framework which serves to cross-identify people, places etc. As long as they play the same role in my personally remembered past, I can treat them as identical [...] even though I do not remember enough of them to say (truly) that I remember who, where, or what they are, and although they therefore are not well-defined individuals by descriptive criteria. (Hintikka 1972, 218–219)

In some chapters of Intentions of Intentionality (1975), Hintikka came close to outlining a counterpart theory when he wrote about the objects in the possible worlds as being "stages", "manifestations", or "embodiments" of the objects in the actual world. This type of account, combined with Hintikka's view on the primacy of cross-identification methods, would have provided a proper neo-Kantian alternative to Kripkean realism, according to which our crossidentification methods would construct the individuals for us while the objects as such would have remained transcendental.⁴⁵ This would have been a plausible view in many epistemic contexts, for instance in the case of perceptual knowledge and perceptual belief. After all, the targets of our perceptions are properties or features that objects possess, not objects themselves. In many respects, the objects themselves are obsolete in the possible world analysis of perceptual attitudes: if we say that a has P in w and v it merely means, in a perceptual context, that a represents some perceived target trackable across modal space and that this target has a further property P in w and v. If I perceive a certain kind of dog, then what matters for the trans-world sameness is that the perceived cluster of properties stays the same from world to world. The perceived cluster of properties may be possessed by Rufus in w, Preston in v, and Duke in z but these dogs may, depending on underlying individuation processes represented by world-lines, be physical/public counterparts or pespectival counterparts of the dog I actually perceive. If the dogs are picked out by a world-line then there are counterpart relations prevailing between the dogs and I have an epistemic attitude directed to a particular individual dog which, as an entity, is mainly constructed by my own

⁴⁵ Hintikka writes about his neo-Kantianism on many occasions, but the exact content of this view remains obscure. At times he stressed the constitutive role of the conceptual arsenal (Hintikka 1969, 172–177) and at times he spoke about perceiving material objects directly (Hintikka 1969, 177–178).

conceptualizations.⁴⁶ This requires elaboration, but we may at least say that at the level of semantic theory, world-lines are partial functions operating on some criteria: if the world-line is perspectival, then the *sameness*, that is, a counterpart relation, is an agent-relative matter and the criteria that the world-line operates on has to do with how an object appears to an agent. If the world-line is public then the counterpart relations also prevail from a public, more objective viewpoint, and the world-line operates on more specific physical or public criteria.

Interestingly, David Lewis (1983) discussed these ideas of Hintikka about individuation in the context of his own counterpart framework and sketched a metaphysical theory of neo-Kantian objects. I shall not go into these details here, but Lewis' views would have provided a starting point for Hintikka if he had been interested in developing neo-Kantian metaphysics involving counterparts to ground his world-line semantics. But Hintikka did not elaborate on his metaphysical presumptions. He never clearly stated what his "semantic neo-Kantianism" means exactly. Instead of making the modal neo-Kantian claim explicit, namely that the individual a in w and the individual a in v are never identical but at most counterparts, Hintikka repeatedly wrote about the sameness of denoted objects across modal space. My suggestion is that we drop the jargon of "uniqueness of reference" and "well-defined individual" that allude to the objective cross-world identity and attribute to Hintikka the view according to which objects in the domains of worlds are, at best, some sort of counterparts to each other. Cross-world identity simply does not occur in Hintikka's setting and the question of when one specific individual in one possible world is identical with another individual in another possible world is not sensible, or at least the answer is always negative. This view does not solve all the problems concerning, for instance, knowing who -constructions or the question of when substitutions of codesignative terms are legitimate but it does clarify Hintikka's position concerning rigidity and the basic design of modal semantics: Hintikka is to be interpreted as an opponent of the standard Kripkean design which posits, first, objects that belong to the domains of a number of worlds and, second, means to track those objects from one world to another, such as constant variable assignment functions.

As the result, this neo-Kantian interpretation of the world-line approach does not support linguistic, semantic, or formal rigidity. However, we may accept the outline of Niiniluoto's account of quantification above, rely on two types of world-lines (which admittedly require more elaboration), and more or less ignore the objects themselves occupying possible worlds.

It is also possible to represent an epistemic variant of semantical rigidity which is significantly weaker than the one defined above but which in certain limited

⁴⁶ One of Hintikka's (1975, 30–31) formulations, namely his doctrine that individuals are world-lines, comes close to this view. Also, Tulenheimo's (2009) reconstructs Hintikka's view along these lines. However, both Hintikka and Tulenheimo (2009, 390) held on to the ideas of "uniqueness of reference" and "well-defined individual", and explained the importance of formulae such as $\exists x K_a(x=b)$ by appealing to those ideas. My suggestion is to abandon the rhetoric and requirements of "uniqueness" altogether. In a more recent book Tulenheimo (2017) departs from these ideas and develops a semantic framework that may perhaps be described as "Hintikka-inspired" since its central tenets appear to be quite independent of Hintikka's views.

contexts nevertheless allows the counterfactual variation of accidental properties. The idea is this: suppose we operate in a scientific context and are engaged in classifying liquids. If we have a natural kind predicate with specific externalist satisfaction conditions at our disposal, such as is water as a theoretical term of science, or just W in short, then this predicate is expected to be true only of substances with the inner structure of H₂O. If we now pick out a sample of water and state with a physical quantifier that $\exists x K_a W x$ then the physical world-line involved should not operate on the basis of a traditionally understood descriptive condition, that is, a description of superficial properties. Instead it should, in such an ideal epistemic situation, operate on a well-defined physical criterion and pick out only samples of H₂O from world to world. In such case, the actual sample is not identical to any sample in any other possible world and hence the variable is not rigid, that is, the initial quantifier does not really bind the variable in the scope of K_a in the sense of quantifying-in, but the formula guarantees that whatever is picked out in the possible worlds introduced by the modal operator, it is H₂O. A perspectival quantifier, in turn, would be appropriate if a descriptive condition linked to superficial properties is involved and especially if the context is not scientific. Sometimes we may use the predicate is water not as a theoretical term but instead to speak about any watery stuff on the basis of our perceptual impression, and in such a case the world-line involved is expected to pick out any samples that have the typical accidental surface features of water. The physical world-line tied to externalist satisfaction conditions of some theoretical term is thus the closest thing to semantical rigidity one may have in the approach at hand.

Before moving on, let me point out that world-lines do not solve the problem of duality relations of quantifiers discussed in section 2.10. Hintikka (1969, 96) explains why co-designative terms are not always interchangeable in epistemic contexts by appealing to his old ideas as presented in $K \not\subset B$, namely to the ideas of contingent identity and "splitting" individuals. If we allow models in which individuals "split" in the sense described in section 2.10, then $\forall x A$ is not generally equivalent to $\neg \exists x \neg A$. The model described in 2.10 makes this clear. This time we shall consider an epistemic interpretation of the model which was arguably Hintikka's original intent. Consider the contingent identity/counterpart model at the end of section 2.10, in which one individual b exists in the possible world b0 while two of its counterparts, b1 and b2, exist in b3, and b4 is a one-placed predicate only true of b3 in b4.

Assuming that the world-lines "split" or "dissect" as, according to Hintikka, they sometimes do, the quantified formulae $\exists x P_a F x$ and $\exists x \neg P_a F x$ are both true in w. One world-line associated with b picks out d at v^* , making $\exists x P_a F x$ true in w, while another world-line picks out c in v^* , making $\exists x \neg P_a F x$ true in w. But since b is the only individual in w the universally quantified formula $\forall x P_a F x$ is also true in w. Thus the duality of quantifiers breaks down. Now, Hintikka states that individuals should not "split" in the case of physical/public quantification and perhaps, according to him, such models should not emerge in the logic of coherent epistemic attitudes. But we should, nevertheless, be able to describe these

situations coherently. Especially in the case of perspectival individuation (which allows "splitting"), we should be able to reason with quantifiers. But the breakdown of the duality relations of quantifiers considerably complicates and even distorts Hintikka's world-line logic. However, I shall not go further into this question here.

3.5 GAME-THEORETICAL SEMANTICS

Hintikka presented his final views on rigidity in his book *Socratic Epistemology* (2007), which basically repeats the central points already made in Hintikka & Sandu (1995). These views are not compatible with the neo-Kantian interpretation presented above, and not compatible with Hintikka's view on the importance of a proper cross-identification system.

In the early 90s Hintikka launched his "second-generation epistemic logic" governed by game-theoretical semantics (GTS), and along with that change of paradigm he acquired an objectual interpretation of quantifiers (Hintikka & Sandu 1995, 249). Roughly, GTS associates to each non-atomic formula of a given interpreted language a two-player semantic game. The two players may be referred to as *Eloise* and *Abelard* according to their game tools, that is, the existential and the universal quantifier, or as *the verifier* and *the falsifier*, according to their respective aims in the game. All non-atomic formulae are reinterpreted as quantificational formulae that can be decomposed by applying some of the game rules concerning the quantifiers.⁴⁷ The game ends when a player confronts an atomic formula. If the atomic formula is true then Eloise wins, if false Abelard wins. Ultimately the truth and the falsity of a non-atomic formula depend on whether there exists a winning strategy for one of the two players. I do not go further into the details of GTS. The specifics, the foundations, and the state of the art may be studied in Sandu (2015).

Regarding rigidity, the aim in Hintikka & Sandu (1995) was to show that rigidity is a nonexistent phenomenon (281) by presenting a long argument according to which, roughly, (i) rigid designation and the substitutional interpretation of quantifiers depend on each other; (ii) substitutional interpretation does not work for quantifiers in second-generation epistemic logic; therefore (iii) rigid designation is dispensable in second-generation epistemic logic (252–258). The plausibility of this argument is not relevant here. If we concentrate on rigidity, the semantical clauses for modal operators and quantifiers given (249) are objectual and in line with the standard Kripkean treatment of variables: the

⁴⁷ GTS regards $A \wedge B$ as a universally quantified formula stating roughly that "Every one of the sentences A, B holds". In a semantic game $for A \wedge B$ the player Abelard or the falsifier chooses whether the game should proceed for A or for B. Similarly, the disjunctions become existentially quantified formulae about sets of sentences stating that some of them hold, and it is for the player Eloise or the verifier to choose with which disjunct the game proceeds. In the case of quantifiers themselves, the game for $\forall xFx$ proceeds as follows: Abelard chooses an object from the domain as the value of the variable, say a, and the game proceeds for Fa. The procedure is the same with the existentially quantified formulae, except that Eloise chooses first. Negation amounts to a switch of roles between the players.

variables come out as rigid when quantified-in, due to assumed constant variable assignment functions. The authors also explicitly recognized this fact: according to their explanation, formulae such as $\exists x K_a F x$ state "that something is true (viz. that F x) of one and the same individual a in a range of different possibilities". Hence, Hintikka ended up abandoning his neo-Kantian orientation after all and explicitly accepted the assumptions that, first, it makes sense to speak about objects as being the same from world to world and, second, that a semantic theory should possess the means to track those objects from one world to another. The paper also repeats Hintikka's old points concerning world-lines and the importance of crossidentification system.

Two observations: first, Hintikka's new semantic clauses for quantified formulae are not compatible with his views on the importance of a cross-identification system. Here the same point can be made as above: if we quantify objectually into an epistemic context $\exists x K_a(x=b)$ with respect to a standard Kripke model, then neither properties nor relations considered, nor any identification method employed, by the epistemic subject are relevant to the truth or the falsity of $\exists x K_a(x=b)$ in the sense Hintikka suggested earlier. The only relevant thing is whether b itself, regardless of its observable properties, is assigned as the value of x in every relevant possible world.

Second, if we accept these assumptions then there is no room for such a sweeping criticism of rigidity that Hintikka's earlier writings seem to suggest. Hintikka is still in a position to deny linguistic rigidity, that is, he may coherently deny the rigidity of any type of natural language expression, including proper names. But by accepting semantical rigidity at the level of variables, the most radical and the most interesting neo-Kantian features of his rigidity critique are lost. The declaration that there are quantifiers tracking objects themselves from world to world locates Hintikka's position close to the widescopers' position, which, at the end of the day, may be a viable position. But it is a conventional position in contrast to the neo-Kantian, one according to which the above assumptions concerning objects themselves should be rejected.

3.6 INDIVIDUALS AND OBJECTS

This chapter surveyed Hintikka's views on rigidity from the late 50s to this millennium and argued that many of Hintikka's ambivalent remarks concerning rigidity become more comprehensible if, first, three different variants of rigidity are acknowledged and, second, Hintikka's largely implicit doctrine of semantic neo-Kantianism is made more explicit. I also argued that, despite Hintikka's declarations indicating otherwise, semantical rigidity at the level of variables was needed for many applications in his semantic frameworks. However, I showed that Hintikka's two early approaches, namely his model set semantics and his world-line semantics, did not support semantical rigidity (or any other variant of rigidity). I suggested a neo-Kantian interpretation of Hintikka's world-line semantics and

sketched a reconstruction of a weak epistemic version of semantical rigidity by its means. I argued that if one is after a radical rigidity critique then the neo-Kantian interpretation is a viable option. I closed with remarks indicating that Hintikka did not, after all, favour the neo-Kantian interpretation in his game-theoretical semantics. Instead, Hintikka ended up accepting a straightforward semantical rigidity that, first, undermines his most radical criticisms of rigidity and, second, overrides his view that a cross-identification system is indispensable in modal semantics, that is, that modal talk of individuals makes sense only when embedded in a proper cross-identification system understood in terms of world-lines. Hintikka's final view is compatible with Kripke's realism, according to which the individuals that our modal language talks about are the objects surrounding us. The possibility of my desk being green is represented by a model with an accessible possible world such that, first, my very desk as an individual is included in its domain and, second, my desk is green.

This contrasts sharply with Hintikka's world-line approach: first, my desk is only found in the domain of this world. The domains of other possible worlds may contain, at best, counterparts of my desk. Second, my desk is a cluster of properties and as an individual it is a trans-world entity constructed by my individuation processes as represented by world-lines. Therefore there must be a distinction between the individuals that our modal discourse concerns and the world-bound objects in possible worlds. True individuals that our epistemic attitudes concern are, or are represented by, world-lines.

4 WORLDS

This chapter assesses worlds of epistemic space by studying the notion of *perspectivity* that we encountered in the previous chapter. Many of our beliefs and other epistemic states are perspectival: they represent the world as being a certain way with respect to ourselves. Perspectival epistemic states provide us a test case for further study of quantification in epistemic contexts.

Hintikka's discussions of perspectivity were motivated by the question of cognitive significance, familiar from the debates concerning propositional attitudes. More accurately, Hintikka introduced perspectival quantification in order to make his modal semantics even more sensitive to cognitive significance. In the current debates it is widely agreed that co-designative terms such as Dr. Jekyll and Mr. Hyde are not generally interchangeable in epistemic or propositional attitude contexts due to a cognitive factor such as a way of thinking about an individual: believing that Mr. Hyde is a murderer involves one way of thinking about the individual, while believing that Dr. Jekyll is a murderer involves another way of thinking about the same individual, and for that reason the terms are not interchangeable without changing the topic, that is, the epistemic attitude under discussion. The debates concern not the existence or the relevance of this underlying cognitive factor but rather its place in the semantics/pragmatics division.

Kripke's model-theoretical semantics and especially Kripke's (1980) later emphasis on rigidity, direct reference, and metaphysical modality have inspired a number of current views according to which cognitive significance has no place in the semantics of propositional attitudes. 48 Model-theoretical semantics in general is not very well suited to accommodating cognitive significance - one might even say that its central tenet is to ignore cognitive aspects of meaning. The so-called neo-Russellian view, for instance, generalizes this tenet by claiming that, despite its pragmatic impact, cognitive significance has no place in semantics and therefore co-designative terms are interchangeable across the board as far as semantics is concerned. It is clear that such a view does not suit Hintikka's purposes. According to Hintikka, it is crucial that our semantics for epistemic logic makes a distinction between knowing that Mr. Hyde is a murderer and knowing that Dr. Jekyll is a murderer. The notion of knowledge that Hintikka's epistemic logic aims to model simply imposes this distinction on us. Hence, from a neo-Russellian point of view, Hintikka's basic units of epistemic reasoning are enriched with pragmatics as they involve agents' intentions or how they think about individuals right at the outset. From a viewpoint of epistemic logic, all these aspects are included in its semantics.

.

⁴⁸ This is not Kripke's own view. Kripke has stated on many occasions that his semantic views concerning statements expressing necessity and possibility are not directly applicable to statements expressing epistemic attitudes.

In the direct reference program it has been customary to separate cognitive significance from semantics in a neo-Russellian fashion. There has been, however, a recent shift of paradigm. For instance, direct referentialists Kaplan (2012) and Recanati (2012; 2013) have argued that cognitive significance should be taken into account in terms of ways of having in mind (Kaplan) or mental files (Recanati) within the framework of model-theoretical semantics.

Cognitive significance is not foreign to semantics. For the maximum explanatory power, our semantic theory should countenance cognitive content, objective content, and extensions. (Kaplan 2012, 141)

The basic test case for this kind of semantics is expressions with context-dependent content, especially perspectival or *de se* content, which present a spectrum of problems of their own. For instance, according to Kaplan and Recanati, two occurrences of the pronoun *I* may come with the same Kaplanian character or with what Recanati calls a "linguistic mode of presentation" yet they may differ with respect to ways of having in mind or mental EGO-files that represent perspectives in individual psychologies. Characters and linguistic modes of presentation contribute to objective content while ways of having in mind and EGO-files contribute to cognitive content. These considerations are, of course, inspired by Frege's work and in this particular case of *I* by his view that we all have a unique intimate perspective on ourselves.

However, perspectival contents undermine the most economic accounts of communication, agreement, and disagreement, since these are taken to require contents that may be shared. Perspectival content is not shareable, at least in cases where agents do not share a perspective. Chalmers (2002a) has argued that the situation calls for a specific two-dimensional view of modal content with a Kripkean metaphysical dimension for objective content and another epistemic dimension for cognitive significance and perspectival content. Chalmers' framework has interesting while unnoticed points of contact with both Hintikka's model set approach and world-line semantics.

Thereby the comparative point of view acquired in the previous chapters will be adjusted in this chapter to compare Hintikka's and Chalmers' views. This leads us, eventually, to the question of how we should understand the worlds of epistemic space: are they epistemically possible or metaphysically possible or a combination thereof?

4.1 PERSPECTIVAL CONTENT

In Plato's *Euthydemus* two *pankratiasts*, "all-round sportsmen", argue that it is impossible for two people to contradict one another (285d–286b). If I describe a man as wounded and you describe him as not wounded, then we are talking about two different things: I am talking about a wounded man and you are talking about a man who is just fine. If you are right then the wounded man does not exists, and I am not talking about anything at all. Hence I cannot contradict you because we are either talking about different things or one of us is saying nothing. This small sophism of two ancient sportsmen is still acutely relevant to the contemporary debate over the nature of content.

Perhaps the simplest view of content is the aforementioned neo-Russellian view that depicts contents expressible with directly referential expressions with objects themselves as their semantic values. The objects form singular propositions that serve as contents of certain sentences, and arguably, as contents of certain beliefs, namely singular thoughts. If I think about a particular wounded man, then the content of my thought is the wounded man. If the wounded man does not exist then my thought has no content, as Plato's sportsmen argued.

The neo-Russellian notion of content has been very fruitful, but it has been argued that, from a wider point of view, the account is too sparse. If *understanding* involves knowledge of content, and if *translation* involves preservation of content, then what is usually understood and preserved does not seem to be a neo-Russellian content. For instance, it has been argued that speakers with no scientific education may understand, at least partly, sentences (and evaluate the correctness of their translations) concerning plutonium, silicon, ethanol, and what not, while being quite ignorant about these substances.

It has been also argued that many of our beliefs are "de se" (Lewis 1979), "centered" (Lewis 1979), "egocentric" (Jackson 2004), or "perspectival" (Kölbel 2018, Sandgren 2018, Recanati 2007, Hintikka 1969; 1975) with respect to their contents, and that these kinds of contents are not neo-Russellian contents. This has been argued convincingly by, for instance, Perry (1979), Lewis (1979), and many others, albeit some disagree (see e.g. Stalnaker 1981, 2008, and Cappelen & Dever 2013). I will not assess this debate here. I presuppose that, first, many of our beliefs are perspectival as they represent the world as being a certain way with respect to ourselves, and that, second, these beliefs cannot be fully modeled as relations to a standard sort of content, that is, to a proposition understood as a set of standard metaphysically possible worlds.⁴⁹

It is also widely held that perspectival contents are problematic in connection with communication, agreement, and disagreement, since these are taken to

⁴⁹ All the participants of the discussion mentioned so far presuppose that, in general, contents may be characterized within a possible world framework. This, of course, is the main presupposition of this chapter.

require more objective, shareable contents.⁵⁰ Perspectival content is often not shareable, at least in cases where subjects do not share a perspective.

Some proponents of two-dimensional semantics (2D) hold that the situation requires for the recognition of more than one kind of content. For instance, Jackson (2004) and Chalmers (2011a) have argued that *primary content* (or *A-intension*) is a perspectival dimension of content, captured by a set of centered worlds, or "worlds considered as actual", while *secondary content* (or *C-intension*) is similar to neo-Russellian content. According to Chalmers (2002a), communication, agreement, and disagreement are to be explained by appealing to the interplay of the two dimensions of content. Roughly, secondary content provides objective shareable content that may be communicated and agreed/disagreed on while primary content has to do with subjects' perspectives on the matters communicated.

After setting the stage in sections 4.2, 4.3, and 4.4 below I evaluate Chalmers' account in section 4.5. First, I try to explain how this rather complicated view is supposed to work by navigating through the jungle of details and prima facie complications that the account is bound to face. Then in section 4.6 I briefly discuss a further interesting problem pointed out by Alexander Sandgren (2018). In section 4.7 I suggest that an account inspired by Hintikka's work offers some unrecognized prospects in this connection. Instead of primary and secondary dimensions, this account appeals to the perspectival and physical/public dimension of content familiar from the previous chapter and then proceeds to explain the aforementioned phenomena in the same fashion as Chalmers' account. From a systematic point of view, my conclusion is that the Hintikka-inspired framework offers some prospects in this area, especially for those who prefer a unified account of perspectivity involved in communication, agreement, disagreement, same-saying, joint wishes, shared interests, etc. From a historical point of view, I take it that the applicability of Hintikka's modes of individuation to the aforementioned questions shows that Hintikka anticipated some current views on modal dimensions.

Before proceeding, I point out that many participants of the discussion, for instance Jackson (2004), Weber (2013), and Sandgren (2018), emphasize that their arguments concern *mental content*, not *linguistic content*. I take this as my starting point as well, albeit the view I set forth may end up positing a more intimate relationship between the two than most current accounts.

I also take note that one may find many examples of different kinds of perspectives in the literature: contents, both linguistic and mental, may plausibly be perspectival in many ways, and these ways probably do not reduce to a presence of a single factor such as indexicals in language or indexicality of thought (see e.g. articles collected in García-Carpintero & Torre 2016, and Recanati 2007). Here I concentrate on two examples: "essentially" perspectival content in the classic sense

 $^{^{50}}$ More accurately: perspectival contents are problematic in all contexts in which shareable contents are required.

of Frege (1918/1956) and Perry (1979), and perspectives on natural kinds, as they are depicted for instance in Schiffer (1987, Ch. 3).

4.2 TWO VARIANTS OF PERSPECTIVAL CONTENT

Frege's famous passage in "The Thought" (1918/1956) sets the stage for the contemporary discussions about perspectival content:

Now everyone is presented to himself in a special and primitive way, in which he is presented to no one else. So, when Dr. Lauben has the thought that he was wounded, he will probably be basing it on this primitive way in which he is presented to himself. And only Dr. Lauben himself can grasp thoughts specified in this way. (Frege 1918/1956)

Perry (1979) continued from this and provided many examples that may be taken to establish that there are irreducibly perspectival contents such as those that Frege alluded to: there are thoughts involving Dr. Lauben that only Dr. Lauben himself may grasp. But Perry was not happy with this, or with what he called "limited accessibility" (1979, 15–16), namely the claim that there are irreducibly perspectival contents that only privileged subjects may grasp. Perry (1979, 16) wrote that he did not believe "in a universe that has, in addition to our common world, myriads of private perspectives". However, Perry's examples and his considerations of "relativized propositions" (1979, 13–15) may be taken to support Fregean limited accessibility. This has generated a vast literature in which perspectival content with limited accessibility has remained a focus (see e.g. papers in García-Carpintero & Torre 2016). The Hintikka-inspired framework introduced below is neutral with respect to the limited accessibility thesis: the framework is suited to capturing privately accessible perspectives but one may also operate with generally accessible perspectives, whatever they may be according to one's preferred ontology.

The Frege-Perry case is an example of a perspective through which one is given to oneself in a unique, perhaps limitedly accessible way. I also wish to consider also another case, namely a perspective through which one may be acquainted with a natural kind (perhaps in a generally accessible way, but still perspectivally). I discuss this example here (especially in section 4.5) because it is simpler than the Frege-Perry case and it allows us to assess some of the relevant issues, for instance descriptivism and the nature of worlds, without directly addressing the difficult question of limited accessibility.

Schiffer (1987, Ch. 3) argues plausibly that beliefs concerning natural kinds are often formed with concepts or predicates from a subject's perspective, and that these do not coincide with any "objective" concepts or predicates that may occur in "objective" beliefs. Schiffer's whole argument goes against the view that beliefs are, in general, relations to propositions. We are interested, for the time being, only

in perspectives, so it suffices for us to merely consider some of Schiffer's examples. One of them involves a belief such as Rusky is a dog. If Tanya believes that Rusky is a dog, then she believes that Rusky possesses a certain property. What is this property? It has been suggested that it is a property of belonging to a certain natural kind.⁵¹ Constitutive of this natural kind is assumed to be a certain genotype, say XYZ, described in biochemical terms. However, the content of Tanya's belief that Rusky is a dog is not that Rusky belongs to this kind whose members possess the genotype XYZ, since Tanya may believe the former fact but not the latter (Schiffer 1987, 57-58). Apparently, someone engaged in biochemical study might have it the other way around: someone might believe that Rusky belongs to XYZ-kind without believing that Rusky is a dog. Now, we may hope that our beliefs involving natural kind concepts have something to do with scientific descriptions of those kinds. If they really do, it cannot be the whole story. Natural kinds enter our beliefs under modes of representation, stereotypes, or some other auxiliary conceptual components, which have to do with how we conceive those kinds. But the problem with a mode of presentation, individual concept, stereotype, or other such item, say the P, is the same as with XYZ-kind. Someone may believe that Rusky is a dog but not that Rusky instantiates the P, and vice versa. 52 Apparently the P should belong to a certain special class C of modes of presentation or individual concepts that are guaranteed to provide the members of Canis familiaris as the targets of beliefs about dogs. Then we should specify C.

This may not be the easiest of tasks in view of the fact that Irish wolfhounds, German shepherds, poodles, chihuahuas and dachshunds are all dogs, while timber wolves, coyotes, and jackals are not, and in view of the fact that in Tanya's place we may have any one of the following: Helen Keller; a man who, though he has seen a few dogs, is as ignorant about them as Putnam is about elms and birches; a person who has never encountered a dog but has read about them; a child who has no biological sophistication whatever and would obdurately persist in calling a Twin-Earth doglike non-dog a dog even after having been apprised of the creature's genetic dissimilarity to our dogs. (Schiffer 1987, 66)⁵³

Hence beliefs about natural kinds are problematic in many ways and, most importantly for the present topic, they involve perspectives.

⁵¹ Some argue that biological species are not natural kinds in this sense. If so then we may alter the example and consider, for instance, an element.

⁵² The P may contain also "indexical" elements, as it is sometimes suggested, such as the same kind as those creatures over there, but that does not solve the fundamental problem, see e.g. Schiffer (1987, 66–67) and Sainsbury (2014, 3–5).

⁵³ See also Sainsbury (2014, 3–5) for similar remarks.

4.3 LIMITED ACCESSIBILITY

Let us return to Frege's original case. Frege points out that Dr. Lauben may want to communicate with others his thought that he has been wounded; but "he cannot communicate a thought he alone can grasp" (Frege 1918/1956, 333). Communication, among other things, requires shareable contents, and perspectival contents with limited accessibility are by definition not shareable. We have a sweeping problem when we add that

It seems that most of what we communicate involves claims about ourselves, claims about our environment, claims about how things currently are, were or will be, claims about others around us. In our ordinary verbal and written communication, personal pronouns, temporal and spatial indexicals abound. Upon reflection it seems that very rarely do we communicate information about what the world is like without any regard for our perspective within it. (Torre 2013, 11)

Besides communication, any activities that require shareable contents become impossible in the presence of perspectival contents with limited accessibility. The problem may be further illustrated in the case of communication. According to the most economic models of communication, such as the "Package Delivery Model" (Moss 2012), successful communication proceeds roughly as follows: a speaker has a belief with certain content, say p, and expresses that belief by uttering a sentence whose content is identical to his belief: "p". Then a hearer understands this sentence, that is, she recognizes the content and comes to believe that p (if she trusts the speaker).

Package Delivery Model #1

- 1. SPEAKER believes: p
- 2. SPEAKER says: "p"
- 3. HEARER comes to believe: p

The model rests on two principles: the first, the mind-to-speech principle, states that the content of the utterance is identical to the content of a belief that the speaker expresses; and the second, the speech-to-mind principle, says that the content of the utterance is identical to the content of a belief that the hearer acquires (Weber 2013). But if the speaker's belief is perspectival and its content limitedly accessible, that is, "a thought he alone can grasp", then the speaker may believe p and utter "p" but the hearer is not in a position grasp p. The hearer may grasp something else, perhaps q, which may be in some way related to p. But our simple model is not equipped to take that relation into account. The model requires a package, and if there is no package the model breaks down:

Package Delivery Model #2

- 1. SPEAKER believes: p
- 2. SPEAKER says: "p"
- 3. HEARER comes to believe: *q*

The perspectival belief *p* of the speaker, due to its limited accessibility, is bound to be different with respect to its content from any of the contents that the hearer may recognize, and hence any belief that the hearer may acquire is bound to be different from *p*. In other words, the speech-to-mind principle fails. There is no single package and there is no delivery. All attempts to communicate perspectival content with limited accessibility are unsuccessful according to the simple model. If most of what we communicate involves a perspective with limited accessibility we arrive at the conclusion that there is virtually no successful communication at all. But surely there must be successful communication that involves perspectival contents. When Dr. Lauben was wounded, he no doubt managed to communicate his limitedly accessible thought to his comrades in some form or another. But the Package Delivery Model fails to illuminate processes of communication involving perspectives.

Similar considerations apply to agreement and disagreement: If we say that the speaker and the hearer agree on some perspectival content, then what exactly they are agreeing on? Their perspectival contents are bound to be different due to their differing perspectives with limited accessibility, and hence the speaker agrees that p while the hearer agrees that q. But that is nonsense unless a very specific relationship between p and q has been established. In the same vein, genuine disagreement also seems impossible: due to the perspectives involved, the speaker's and the hearer's seemingly conflicting beliefs may very well be consistent (p) and not-q are not contradictory).

4.4 CENTERED CONTENTS

Lewis (1979), Egan (2006), and Ninan (2013) have suggested a way to capture perspectival contents as sets of ordered pairs consisting of an ordinary possible world and a center, $\langle w, v \rangle$, where a center, in turn, is taken to be, for instance, a pair of an individual and time $\langle i, t \rangle$. The resulting propositional attitude is a *centered indexical thought*, a thought whose complete truth-conditions depend on contextual factors such as who is in the center, that is, on parameters i and t (Recanati 2016). The limited accessibility thesis does not apply fully to this kind of centered content. A centered content may be shared in the sense that Dr. Lauben may think that he is wounded at a certain time a and it is true just in case Dr. Lauben is wounded at a. On the other hand, I may think exactly the same centered indexical thought now, and it is true if I am wounded now (see e.g. Recanati 2016).

The troubles discussed in the previous section nevertheless remain. Now Package Delivery Model #1 predicts that the following exchange is successful:

(i) Dr. Lauben believes that he is wounded (as a centered belief p that the person designated by the center is wounded) and (ii) he communicates this centered belief p to someone, say, Rudolph Lingens but (iii) if Rudolph Lingens acquires p he comes to believe that the person designated by the center, in this case Rudolph Lingens himself, is wounded. This because of the special nature of p as an indexical belief centered around the believer. There is a single package, the centered content p, and it gets delivered through the chain, as illustrated by model #1 above (see Stalnaker 1981). Both of the principles that the model rests on are obeyed but still the result is still, of course, wrong. What we would like to have is a model such as #2, where Rudolph Lingens acquires not the centered content where the thinker designated by the center is wounded but the content that Dr. Lauben is wounded. But this content is not the centered content p but another content, perhaps q, according to which a certain person who is not designated by the center, namely Dr. Lauben, is wounded. According to model #2, this is unsuccessful communication since there is no single package and no delivery. Therefore the Package Delivery Model is too simple to capture communication involving centered contents.

There are many solutions to the problem of communication involving centered contents, for instance, solutions utilizing multiple centers: Lewis (1983), Ninan (2010), and Torre (2010) include as many individuals in the center as there are conversationalists. The center may be, for instance, (Dr. Lauben, Rudolph Lingens, t), that is, a triple of an individual, a second individual, and a time. Thereby a centered content is argued to be shareable to all conversationalists who are in the center. The surrogate content accounts, such as in Moss (2012) and Kölbel (2013), invoke surrogate non-perspectival shareable contents that may be extracted from the perspectival contents. Gibbard (2012) and Weber (2013), in turn, appeal to an operation that Weber calls "re-centering": the hearer performs an operation that shifts the center from one individual to another. All these solutions have their own problems (see Pagin 2016), but I do not assess them here in detail. I just point out in general that all the centered accounts of communication fall short in explaining perspectivity in other related phenomena that require shareable contents but do not involve a communicative element such as agreement, disagreement, samesaying, etc. We may agree, disagree, or say the same thing, or have a common intentional attitude without communicating it. Dr. Lauben and Rudolph Lingens may, for instance, disagree on whether Dr. Lauben is wounded without communicating their judgments. An intuitive notion of non-centered standard content explains what is going on. There are two parties and a single content, Dr. Lauben is wounded, and the first party believes it while the second party believes its negation. Hence there are contradictory attitudes and a genuine disagreement. But the centered content account offers a picture reminiscent of Plato's dialogue: Dr. Lauben's centered belief that he is wounded and Rudolph Lingens' centered belief that Dr. Lauben is not wounded are, surprisingly, compatible. In the first case the

person designated by the center is wounded, while in the latter case an individual outside the center is not wounded, and these are perfectly compatible different contents. On the other hand, if Dr. Lauben thinks that he is wounded while Rudolph Lingens thinks of himself that he is not wounded, the centered content account predicts disagreement, since these contents are not compatible. In the first case the center is wounded and in the second case the center is not wounded. These results are obviously wrong and here no account of communication is helpful, that is, no account of expression and acquisition of information can solve problems having to do with our judgments of disagreement. Similar considerations may be extended to other notions as mentioned above.

Next, I consider views that posit two dimensions of content. The prospects of these views are, first, that we may pursue a unified account of communication, agreement, and other closely related phenomena involving perspectives and, second, we may allow the existence of irreducibly perspectival contents with limited accessibility without worrying about how to convert them into shareable contents through operations such as "re-centering" or how to extract shareable surrogate contents from them by some other means.

4.5 DIMENSIONS OF CONTENT

The so-called epistemic interpretation of two-dimensional semantics recognizes two dimensions of content designed to, first, illuminate the nature of perspectival beliefs and, second, explain communication, agreement, and disagreement, and other related phenomena when perspectives are involved. The outline of the first dimension or primary content bears some similarity to the centered accounts mentioned above. Roughly, Chalmers (2002a, 611) holds that primary content characterizes a believer's perspective by invoking a set of centered worlds. Jackson (2004, 261) holds that perspectives are captured by A-intensions, which, in turn, invoke sets of "worlds considered as actual". Centered worlds and worlds considered as actual are thereby supposed to help to capture the perspectival contents. I assess the question of how this is supposed to work in greater detail below. Roughly, the speaker's perspectival belief that Dr. Lauben is wounded is captured by the set of centered worlds where the speaker is in the center and Dr. Lauben is related to the center in a way that the speaker thinks Dr. Lauben is, or by the set of worlds that the speaker considers as actual. The hearer's perspectival belief is captured, in turn, by the set of centered worlds with the hearer in the center, and Dr. Lauben is related to the center as the hearer thinks he is, or by the set of worlds considered as actual by the hearer. These sets are typically different sets. If the speaker is Dr. Lauben, then he is in the center and related to himself as he thinks he is, perhaps in a unique and limitedly accessible way. If the hearer is Rudolf Lingens, then Lingens is in the center and Dr. Lauben, a wounded person who is outside the center, is related to Lingens as Lingens thinks Dr. Lauben is. Hence the believed contents, that is, the primary content of the speaker's belief and the primary content of the hearer's belief, are different (Chalmers 2002a, 631, n. 29). Note the apparent descriptive orientation here: Dr. Lauben is individuated by how a speaker or a hearer thinks Dr. Lauben is related to the center, which assumedly happens by associating certain individuating properties with Dr. Lauben. This leaves room for Kripkean *Gödel*-type examples to indicate that the singular term *Dr. Lauben* is not rigid here. Someone else may possess the relevant associated individuating properties and hence there are possible worlds in which someone else, say Prof. Biergarten, is wounded and related, for instance, to the hearer in the particular way that the hearer thinks Dr. Lauben is (in the actual world). These worlds are included in the set that captures the hearer's perspectival belief. The same considerations may apply to the speaker's perspectival belief in the scenarios involving amnesia (see e.g. Perry 2002).

Chalmers (2002a, 631; 2011, 620–621) holds that, in addition to this primary content, there is another, secondary dimension of content that helps to explain communication, agreement, and disagreement. The secondary content is not perspectival. Primary content determines the secondary content by determining the unique target of the belief in the actual world, namely in this case Dr. Lauben himself and no one else. Then the secondary content is captured by the set of possible worlds where Dr. Lauben is the target (and wounded). Despite their differences, both primary contents, the speaker's and the hearer's, happen to determine the same target in the actual world, namely Dr. Lauben, and thereby the beliefs have the same secondary content. This explains, according to Chalmers, the successful communication between the speaker and the hearer who do not share a perspective. The secondary content also serves as a non-perspectival common ground in agreement, disagreement, and in all potential cases where shareable contents are required. The Package Delivery Model may be reinstated. Now the relevant contents appear as pairs in which the first index is the primary content and the second index is the secondary content. The phenomena under assessment here may be explained by appealing to the same secondary content, r.

Package Delivery Model #3 – \primary content, secondary content\

```
    SPEAKER believes: \langle p, r \rangle
    SPEAKER says: "\langle p, r \rangle"
    HEARER comes to believe: \langle q, r \rangle
```

Jackson (2004) does not appeal to secondary content in this connection. Instead he seeks to explain the phenomena by appealing only to A-intensions, that is, by appealing only to perspectival content.

Let me now fill in some details of epistemic 2D. A standard model-theoretic modal framework familiar from the previous chapters recognizes just one dimension of content. Let us acquire here a slightly modernized notation and begin with a frame, $F = \langle W, R_0 \rangle$, consisting of a non-empty set of points of evaluation W and an accessibility relation on that set R_0 . A model, M, is obtained

from a frame by adding a valuation function V that maps each propositional variable (p, q, ...) to a subset of W. The discussion here presupposes the connection between content and truth (the content of belief is understood as the set of possible worlds in which the belief is true). Hence the satisfaction relation, $M, w \models \varphi$, is intuitively understood here as the truth of a formula φ in a model $M = \langle W, R_0, V \rangle$ relativized to a possible world w, and it is defined with the recursive clauses typical to model theory. Here are, for example, the clauses for propositional variables, negation, and the modal operator \square_0 .

```
M, w \models \varphi if and only if w \in V(\varphi)

M, w \models \neg \varphi if and only if not M, w \models \varphi

M, w \models \Box_0 \varphi if and only if M, w^* \models \varphi for all w^* \in W such that wR_0 w^*
```

Now, suppose that there are two points of evaluation, namely w_0 and w_1 , such that the latter is accessible from the former, and p is mapped to both, that is

$$W = \{w_0, w_1\}$$

$$w_0 R_0 w_1$$

$$V(p) = \{w_0, w_1\}$$

Then M, $w_0 \models \Box_0 p$. If we alter the valuation and let $V(p) = \{w_0\}$, that is, p is mapped only to w_0 while w_1 remains accessible from w_0 , then M, $w_0 \models \neg \Box_0 p$.

Now the basic formal ideas of 2D are easy to grasp. Let me first note, however, that Chalmers developed his constantly changing ideas mainly informally and hence we do not know what his preferred way would be among the many possible ways to introduce dimensions in model-theoretic semantics. What follows is one way to do it, employed in many textbooks. It is probably not the way that Chalmers would have had chosen back in the 90s and it is certainly not the way that Chalmers would choose now (see more below). The basic model-theory presented here nevertheless serves well to illustrate the interpretational problems that Chalmers' 2D faces.

In 2D, truth-values are assigned to propositional variables relative to pairs of worlds. One way to carry this out formally is to set conditions for two accessibility relations, for instance

$$\langle x, y \rangle R_1 \langle x^*, y^* \rangle$$
 if and only if $x^* = y^* \langle x, y \rangle R_2 \langle x^*, y^* \rangle$ if and only if $y = y^*$

The set of points of evaluation, or possible worlds, is $W \times W$. The satisfaction relation, M, $\langle w, w^* \rangle \models \varphi$, intuitively the truth of a formula φ in model

$$M = \langle W, R_1, R_2, V \rangle$$

relativized to a pair of possible worlds, is defined with similar clauses as above. The difference is that the points of evaluation now consist of pairs (with a world as an x-index and world as a y-index), and that the resulting models contain two accessibility relations. This makes room for two modal operators that both obey a clause corresponding to the semantic clause for \Box_0 in one-dimensional semantics. Due to the conditions set for their respective accessibility relations, the operators come apart. For instance, if $W = \{w_0, w_1\}$ then we have four points of evaluation: $\langle w_0, w_0 \rangle$, $\langle w_1, w_0 \rangle$, $\langle w_0, w_1 \rangle$, and $\langle w_1, w_1 \rangle$. Now, if $V(p) = \{\langle w_0, w_0 \rangle, \langle w_1, w_0 \rangle\}$ then $M, \langle w_0, w_0 \rangle \models \Box_2 p$ because R_2 provides an access from $\langle w_0, w_0 \rangle$ and $\langle w_1, w_0 \rangle$ in this particular example, and p is mapped to both. On the same occasion, however, $M, \langle w_0, w_0 \rangle \models \neg \Box_1 p$ because R_1 provides an access to pairs with identical x- and y-indeces, namely to $\langle w_0, w_0 \rangle$ and $\langle w_1, w_1 \rangle$, and p is not mapped to the latter. If we alter the valuation and let $V(p) = \{\langle w_0, w_0 \rangle, \langle w_1, w_1 \rangle\}$ then it may be established that $M, \langle w_0, w_0 \rangle \models \neg \Box_2 p$ but $M, \langle w_0, w_0 \rangle \models \Box_1 p$.

These are the relevant formal details for our purposes. Then there is a question of interpreting the formalism further. How do the phenomena discussed in the sections above relate to the clauses just stated? The general idea is familiar: every substantial epistemic attitude corresponds to a division of possibilities. In some possible worlds, the epistemic attitude in question is true while in others it is false (some possibilities may be neutral: the attitude is neither true nor false). The content of the epistemic attitude is the set of pairs of worlds in which the epistemic attitude is true. The main novelty here is that the division of possibilities may be understood either among the set of world-pairs that R₁ provides an access to or among the world-pairs that R₂ provides an access to. The condition for R₂ states that this accessibility relation obtains only between those pairs that share the same y-index. As the y-index is kept constant, it is only the variation of the first index that is relevant to evaluating $\Box_2 p$ and thereby the two-dimensional evaluation of $\square_2 p$ coincides with the one-dimensional evaluation of $\square_0 p$ above. The post-Kripkean understanding is that this may be used to represent a metaphysically necessary intension of p, a sort of content that renders p as true no matter where or under which metaphysically possible circumstances its truth is evaluated $(\neg \Box_2 p)$ is naturally taken to represent the lack of such an intension while $\neg \Box_2 \neg p$ is taken to represent a contingent intension rendering p as true in some possible circumstance). Hence a division among R₂-worlds is a division among metaphysically possible worlds. If the evaluated item is a belief, then the division captures the content of the belief in terms of what metaphysical possibilities it excludes and what it allows. Chalmers and Jackson understand these aspects of metaphysical modality to be the secondary dimension of content while R_1 and its respective modal operator \square_1 are taken to represent the perspectival primary dimension that has to do with epistemic possibilities. A division among R₁-worlds is understood as a division among epistemic possibilities that are ruled out by the subject and those that are not ruled out by the subject. Together the R_1 and R_2 divisions are taken to capture the two-dimensional content. How exactly does this apparatus capture the perspectival content? The condition for R_1 allows access only to those pairs in which the x-index and the y-index are identical. What makes this feature suitable for representing perspectival content? Let us postpone the answer to this question, and introduce an example in order to assess the question properly.

The interpretational problems involved may be illustrated by investigating whether 2D can capture perspectival beliefs concerning natural kinds and whether Package Delivery Model #3 can capture communication involving perspectival beliefs concerning natural kinds. I choose to discuss natural kinds here rather than the Frege-Perry case for simplicity. Let us start with a pair of beliefs that Jackson and Chalmers assess in many connections. The two beliefs are There is water and There is H_2O , which according to both authors may differ with respect to their primary content (A-intension) but share the same secondary content (Cintension).⁵⁴ One way to understand the differing primary contents and the same secondary content of these beliefs is to consider There is water and There is H_2O as approximations of two perspectives on one and the same target. We may presumably have perspectival beliefs concerning water, and there are situations in which we communicate them. In such a case, successful communication should proceed along Package Delivery Model #3: a speaker may have a perspectival belief concerning water as it is related to him as he thinks it is, and as the result of communication a hearer may acquire another perspectival belief concerning water as it is related to her as she thinks water is. Communication is successful just in case it is secured by the same secondary content: If the perspectival belief of the speaker and the perspectival belief of the hearer determine H₂O as the target of both beliefs in the actual world, then the secondary content is the same and communication succeeds (and genuine agreement, disagreement, same-saying, etc. are possible).

Let us ask whether the primary content really differ from the secondary content in the case of beliefs concerning natural kinds, that is, are primary contents really suitable for capturing perspectives? The problem here is that Jackson (2004, 263) and Chalmers (1996, 57) agree with Kripke (1972/1980) that water is in the metaphysical sense necessarily H_2O . This means that the set of metaphysically possible worlds in which *There is water* is true and the set of worlds in which *There is* H_2O is true are the very same set, and thereby at least the secondary content is the same. But where should we look to find differing primary contents? What are the worlds we should consider? The above model-theoretic definitions recognize only one set of worlds, W, which must be the set of metaphysically possible worlds. This is because Chalmers and Jackson agree that the apparatus captures the metaphysical modality when one point of evaluation is kept constant and the other is allowed to vary among the members of W. In particular, Jackson 2011 (but also

⁵⁴ In what follows, we apply 2D to beliefs. But ultimately this makes no difference. Our topic is communication and the Package Delivery Models rest on principles of *mind-to-speech* and *speech-to-mind*. Thereby contents of beliefs are identical to contents of utterances in this particular case.

Chalmers 2011b, fn. 9) has emphasized that there is only one set of worlds, which is the set of metaphysically possible worlds. Jackson (2011, 131) calls this view "One Space-ism". Thus the solution must involve an explanation of how the pairing of worlds and how R_1 manage to make room for evaluation-pairs in which *There is water* and *There is* H_2O differ in truth value. It is crucial that they differ, otherwise the primary contents cannot be different and the 2D framework cannot capture perspectival beliefs concerning natural kinds.

According to Chalmers' earlier work (Chalmers 1996; 2006) the world-pairs that R_1 provides access to are "scenarios" consisting of metaphysically possible worlds. These world-pairs have to do with what is epistemically possible, that is, with possibilities that cannot be ruled out by *a priori* reflection. If so, then the mere pairing or double-indexing of worlds typical to 2D cannott provide the desired difference in perspectival content between *There is water* and *There is* H_2O since, if they are true in the same metaphysically possible worlds, as Jackson and Chalmers agree, then they are also true in the same pairs of worlds no matter how one carries out the pairing. The purpose of the exercise is to locate different sets of world-pairs in which the above beliefs are true, and the mere pairing of metaphysically possible worlds in a 2D framework does not provide such sets if one proceeds from the starting point that the target beliefs are true in the same metaphysically possible worlds to begin with.

Another idea of Chalmers (2006, 586), already mentioned, is that scenarios consist of centered worlds. This formulation does not clarify the situation either. A designated epistemic agent in the center with a time or a place or any novel parameter does not change the fact that *There is water* and *There is* H_2O remain true in the same metaphysically possible worlds, centered or not.

Yet another idea is that the worlds in question are considered "actual" rather than counterfactual (Jackson 2004, 261; Chalmers 2004, 186). This is what the evaluation in pairs that consist of the same world twice over is supposed to represent (that is, evaluation in pairs such as $\langle w_0, w_0 \rangle$, $\langle w_1, w_1 \rangle$, $\langle w_2, w_2 \rangle$...). Now, this is the answer to the question concerning R_1 above. 2D is supposed to capture the epistemic dimension and thereby perspectival beliefs as sets of the above kind of world-world pairs. The idea is somehow to look at each world by its own standards. For instance, if we have a world exactly similar to the actual world, Twin-Earth, with the only exception being that there is no H₂O but XYZ in its oceans and rivers then the Twin-English speakers in some sense seem to state a truth when they say that there is water in Twin-Earth because by *mater* they refer to XYZ. On the other side, however, the fact remains that XYZ is not water, and if we recall our starting point that There is water and There is H_2O are true in the same metaphysically possible worlds, then it seems that even the pairing of worlds with themselves does not provide pairs in which the truth-values of these beliefs would differ. Also, intuitively this seems to be the case: if the target beliefs share the truth-value by virtue of some metaphysical fact, then our ways of "considering" worlds should not change the situation.

In his later writings, Chalmers (2011a; 2011b) has favored a linguistic interpretation of the possibilities relevant to the primary dimension. In his Conscious Mind (1996) he took worlds as pre-linguistic and primitive (1996, 336), but later he adopted a "metaphysically neutral" approach and constructed the worlds, or scenarios, out of linguistic material, namely sentences. The points of evaluation relevant to the primary dimension are now maximal epistemically complete hypotheses, that is, (possibly) infinite conjunctions of sentences describing possible worlds in an idealized language containing "epistemically invariant expressions" (Chalmers 2011b, 75). These expressions are such that "when s is epistemically invariant, then if some possible competent utterance of s is epistemically necessary, all possible competent utterances of s are epistemically necessary" (2011b, 75). What exact expressions fulfill this condition is not clear, but we are told that at least proper names, natural kind terms, and theoretical terms are not epistemically invariant (2011b, 75). Perhaps these epistemically invariant expressions are similar to the expressions needed in "canonical specifications" (see below) of scenarios. Canonical specifications of scenarios must be given in "semantically neutral" terms that are characterized as expressions that are not "Twin-Earthable", that is, expressions that do not change their meaning in Twin-Earth type thought-experiments (Chalmers 2011b, section 4.2). It may be argued that water means H₂O in Earth but in Twin-Earth where oceans are filled with XYZ, water means XYZ. Hence water is not semantically neutral. In general, Chalmers (2004, 199) states that the idealized canonical language and the complete canonical descriptions consist of a language required to describe "PQTI", the conjunction of microphysical truths (Physics), phenomenal truths (Qualia), a "That's all" truth, and certain Indexical truths.

In any case, the hypotheses are reminiscent of Carnap's state-descriptions and especially reminiscent of Hintikka's model sets, familiar from the previous chapters. Now it is clear, at least, that the evaluation-points relevant to the primary dimension and the evaluation-points relevant to the secondary dimension constitute two different sets. The elements of the first set are maximal epistemically complete hypotheses while the elements of the second set are (perhaps primitive) metaphysically possible worlds. Hence Chalmers abandons Jackson's One Space-ism. However, Chalmers (2011b, 80) revives the spirit of One Space-ism in his "Two Space-ism" by invoking the principle of "Metaphysical Plenitude", which states that for every epistemically (linguistically) constructed world there is a corresponding metaphysically constructed world, that is, all hypotheses describe metaphysically possible worlds. The definitions above must be augmented to meet these complicating new factors. Still, *There is water* and *There is* H_2O are true in the same maximal epistemically complete hypotheses if all these hypotheses describe metaphysically possible worlds.

Then Chalmers (2011b) invokes the distinction between verification and satisfaction. The turnstile symbol \models used in the definitions above represents satisfaction, which according to Chalmers is tied to the metaphysical modality. The relevant beliefs must be evaluated not relative to the satisfaction but relative to the

verification relation tied to epistemic modality. A world w is said to verify a sentence s when a canonical specification d of w implies s. The definitions must be augmented again to accommodate these two relations. This move, however, is not a solution to the initial problem but rather its restatement. Let's say we have a world that verifies There is water but not There is H_2O , that is, from a complete canonical specification containing descriptions of all the microphysical truths of that world, we may derive There is water but not There is H_2O . Our question was: how could that be if the world specified is a metaphysically possible world in which water is H_2O ? If d is a complete specification of w, then it tells us that there is water in w, not by using the term water but by using some "nearby" invariant term "in the vicinity" (2011b, 75), and d tells us that there is H₂O in w, not by using H_2O but an invariant term, and it tells that these are identical. If we may derive There is water from d, then we may derive There is H_2O from d, unless it is the case that d tells us with some invariant expressions that water and H₂O are distinct. Then the world would not be metaphysically possible and both One Space-ism and Two Space-ism with Metaphysical Plenitude would be false.

The fundamental problem we have been circling is not the nature of the points of evaluation or the manner of evaluation, but rather that somehow the points relevant to the primary dimension must outnumber the points relevant to the secondary dimension. One approach is to claim that some hypotheses are not metaphysically possible. These hypotheses would not describe metaphysically possible worlds but rather metaphysically impossible worlds. But what are such worlds like, or more precisely, what kind of sentences these hypotheses consist of? The problem is already present in Kripke's view that there are necessary a posteriori truths. A necessary truth is true in every possible world and it is not possible for it to be false. But "a posterior" means, in this context, that the truth in question is possibly false for all we know a priori: there are epistemic possibilities that we must exclude by resorting to empirical methods before we may come to know the truth in question. Now what are these possibilities? An appealing answer that Chalmers (2011b, fn. 9) also seems to suggest is Two Space-ism without Metaphysical Plenitude, that is, that the possible worlds relevant to a priority and the possible worlds relevant to necessity come radically apart. The latter is evaluated relative to metaphysically possible worlds while the first is evaluated relative to conceptually or epistemically possible worlds (or their specifications) that need not be metaphysically possible. These conceptually but not metaphysically possible worlds would considerably outnumber the metaphysically possible worlds.

One problem with this is the following: true, we may coherently narrow down modalities, and decide for instance that regarding, say, *physical* necessity we only consider worlds in which the laws of physics of the actual world hold. The standard view sees different modalities as concentric spheres. A picture that depicts physically possible worlds as a proper subset of metaphysically possible worlds is plausible. But what is the sphere that is larger than metaphysical possibility, that is, what is the more general, conceptual modality that one may narrow down to get the metaphysical modality, and on what grounds does one

proceed with the narrowing? By saying that for the most general conceptual modality we consider all worlds, but for the metaphysical modality we consider only those worlds in which the metaphysical laws of the actual world hold? What are the metaphysical laws of the actual world?

The official Chalmersian answer is that the more general modality is epistemic: the possibilities that one cannot rule out by a priori reasoning. Apparently it is not a priori knowable that if there is water then there is H₂O. Therefore, among the possibilities relevant to the primary dimension, there should be hypotheses according to which there is water but there is not H₂O. But what does that mean? What kind of world is the one described by such a hypothesis? Kripke's analysis, accepted by Chalmers, has it that we know a priori that everything is necessarily identical to itself, including water. So are these worlds in which water is not H₂O but, say, XYZ? In such a world there is no H₂O. But if one thinks about it, there is no water either. There is only XYZ and XYZ is not water. Hence those worlds do not seem to satisfy nor verify There is water or There is H_2O . Are these worlds such that the water in them is not identical to itself? A world in which water is not identical to itself is not only metaphysically impossible, it is also conceptually impossible: any given thing is identical to itself, and this we know a priori merely by resorting to rational reflection. Many have argued that the distinction between conceptually and metaphysically possible worlds does not make sense, and for more detailed arguments, see e.g. Stalnaker (2003a) and Jackson (2011).

One related intuition involved here is that There is water and There is H_2O differ with respect to their primary intension because the identity claim Water is H_2O is epistemically contingent: beliefs that water is H₂O and that Hesperus is Phosphorus are not knowable to be true a priori and thereby it must be epistemically possible that they are false, that is, there must be epistemically/doxastically possible worlds in which Water is H_2O is false and consequently There is water may be true while There is H₂O may be false, despite the shared Kripkean conviction that all true identities are metaphysically necessary. Kripke's examples of a posteriori necessary truths in Naming and Necessity are usually mentioned in this connection (see e.g. Weatherson & Egan 2011). But as Soames (2011, 87-88) has argued, it is by no means clear that the propositions that Hesperus is Phosphorus or that water is H₂O come out as a posteriori in Kripke's framework: if these propositions are understood as singular propositions then these propositions are perfectly on par with the propositions that Hesperus is Hesperus and that water is water. These propositions merely predicate self-identity to Venus and to water, and the self-identity of each thing is knowable a priori. There are no worlds, epistemically or metaphysically possible, in which the propositions would be false. According to Soames it is rather that the related descriptive beliefs require empirical justification, and it is debatable whether this requirement transfers to knowing the singular propositions.

However, especially in his recent work, Chalmers (2011b, 74) tends to take the notion of epistemic possibility as primitive without elaborating on it. But what exactly is the epistemic possibility involved? There are at least five immediate

candidates for the epistemic possibility that water is not H₂O: first, the possibility that the singular proposition expressed by the sentence Water is H₂O involving rigid designators is false, that is, the possibility that water is not identical to itself. Second, the possibility that a descriptive thought expressed by the sentence Water is H_2O is false, that is, the possibility that the entity that satisfies the description of identifying properties that we associate with water is not the same as the entity that satisfies the description of identifying properties we associate with H₂O. Third, the possibility (not to be confused with the first one above) that the sentence Water is H₂O does not express a proposition involving water/H₂O but some other entirely different proposition that is false. Fourth, the possibility that the substance-names water and H_2O are not co-designative. Fifth, the possibility that the predicates is water and is H₂O are not co-extensional, that is, the satisfaction conditions of the predicates are different. These are quite different possibilities. We have good reasons to think that the first is not a possibility of any kind. It is not metaphysically possible and it is not conceptually or logically or epistemically or doxastically possible that water is not identical to itself. The second possibility definitely exists: it is the plain old Kripkean metaphysical possibility that the greatest student of Plato was not the teacher of Alexander or that the inventor of bifocals was not the first Postmaster General of the United States. Locating such possibilities does not require fancy 2D or other possibilities in addition to the metaphysical. The last three do not necessarily have anything to do with water, since they are all meta-linguistic possibilities concerning linguistic expressions. All are metaphysical possibilities: sentences do not necessarily express what they actually express, proper names do not necessarily refer to what they actually refer to, and predicates do not necessarily have the satisfaction conditions that they actually have or denote the properties they actually do.

Chalmers' 2D is an impressive theoretical framework built to accommodate possibilities that somehow differ from the metaphysical possibilities. But it is not clear what is meant by these possibilities. All we are given is a vague characterization that they are possibilities that cannot be ruled out *a priori*, and then some examples such as *Water is not* H_2O , about which we are told that they may not be metaphysically possible but they must nevertheless be epistemically possible since "there is no possible mental life that starts from that thought [*Water is not* H_2O] and leads to an *a priori* justified [... rejection] of that thought" (Chalmers 2011b, section 3). But what is that thought we invoke when we contemplate the possibility that water is not H_2O ? Is it one of the five variants above, or something entirely different?

Another problem for the potential view that accepts Two Space-ism but abandons Metaphysical Plenitude is that if the primary dimension is associated with one set of hypotheses and the secondary dimension with another set of metaphysically possible worlds then primary content and secondary content can never be the same; they trivially involve different sets of possibilities, and the first set is always immensely larger than the second. However, it is vital for Chalmers that in some cases (for instance when an agent has true beliefs) the two unite or

align such that the worlds described by the hypotheses comprising the primary content are the very same metaphysically possible worlds that capture the secondary content. Chalmers also utilizes a more intimate connection of the two dimensions, for instance in his zombie argument, which contains a step that goes through only if the primary possibility of the most crucial premise entails its secondary possibility (Chalmers 2010, 151). If there are immensely more primary possibilities than secondary possibilities then, of course, the premise may well be primarily possible without being secondarily possible.⁵⁵

Hence it seems that the only plausible answer to this fundamental world-set problem is that the relevant "impossible worlds" are merely metaphysically possible worlds (or sets of sentences that describe them) in which rivers and lakes and taps are not filled with H₂O but some other substance. This is of course perfectly plausible but, first, it is metaphysically possible in the plain old sense, and we may ask why do we need 2D-consideratations to begin with; second, in connection with the application at hand, it is intuitively odd: if we analyze the content of a belief concerning water, why include worlds with no water in them? Even if we aim to capture the perspectival primary dimension of a belief about water, it seems to be a confusion to claim that worlds that contain no water are also relevant to capturing that content. The perspective is supposed to be on water, not on some other stuff. At the level of our formal model theory, this means that if p represents a belief about water then the valuation function relevant to the primary dimension also maps p to some worlds that contain no water, that is, p is true in some waterless worlds. Perhaps a perspective on water may be the same as the perspective on some other stuff. In such cases the framework seems more intuitive.

The linguistic interpretation of worlds does not save Chalmers from this complication. We may illustrate the situation in the previously established framework. First we must introduce some semantic machinery of predicate logic that allow us to move to sub-sentential level: a non-empty domain of objects D together with a domain function that assigns to each $w \in W$ a set of objects, that is, a subset D_w of D which intuitively is the set of objects existing in that world, and an interpretation function I for non-logical vocabulary such that, for instance, for a constant a and a world w,

$$I(a, w) \in D_w$$

 $^{^{55}}$ The premise is (F and not-Q) where F is a complete physical description of reality, possibly an infinite conjunction of microphysical facts, and Q is a fact regarding consciousness, e.g. that someone is conscious or has a certain phenomenal property. The zombie argument rests on the claim that if (F and not-Q) is primarily possible then it is secondarily possible, that is, if there are worlds relevant to the primary dimension in which the premise is true then there are worlds relevant to the secondary dimension in which the premise is true. But if the world-sets relevant to the different dimensions are disjoint or if the worlds relevant to the primary dimension outnumber the worlds relevant to the secondary dimension then, in general, primary possibility does not guarantee secondary possibility. Perhaps the argument is not meant to rest on 2D alone but Chalmers (2010, 151; 1996, 133) analyzes the problem as a gap between 1-possibility and 2-possibility and states that the gap closes in cases when the 1-intension and the 2-intension of the premise are the same. On the other side, Chalmers (2010, 153) also states that for the zombie argument it is not necessary that the 1- and 2-intensions are the same.

The rest of the details may be omitted here. If we try to introduce some sort of perspective to the interpretation of water then, first, the term must be considered as designating different things in different worlds. For instance, suppose that the constant water at w_0 designates H_2O as it does in the actual world but at w_1 it designates not water but some other thing, an element of D_{w1} , say, the Mariinsky Theater in St. Petersburg. Suppose also that the interpretation function does not determine any designation for the constant H_2O at w_1 . Now, whatever our exact semantics and our definition of valuation may be, it seems plausible that it should depict the belief There is H_2O as being false (or at least not true) in w_1 but There is water as being true if we consider water from the perspective of that world as designating the Mariinsky Theater. This kind of perspective may be introduced by letting R_1 open access to the pairs consisting of the worlds themselves twice over while letting R₂ open access to the pairs in which the second component is kept constant. Thereby the truth-value of the belief There is water in the R₁-accessible pairs depends on the designation of water in each particular world under evaluation, while in the R_2 -accessible pairs it depends on the designation water in the world that is kept constant.

The problem with this interpretation from the viewpoint of Chalmers is that it does not relate in any obvious way to a perspective that an agent may have towards water. We have arrived at what is sometimes called the "meta-semantic" interpretation of 2D framework (Stalnaker 2004). It provides possible worlds where beliefs concerning water and H₂O come apart because the term *water* does not designate H₂O in those possibilities. This approach hardly isolates a perspectival aspect of content. Rather, it spells out the fact that linguistic expressions might be used to designate different things than they are actually used to designate, in other words, linguistic expressions may have different contents than they actually have. Applied to our case of beliefs, this interpretation would spell out the fact that our beliefs might have different contents than they actually have.

Chalmers aims for interesting conclusions concerning the perspectival content. In order to reach such conclusions, he must deny the meta-semantic interpretation that represents changes in content from one point of evaluation to another. He needs to state that the content remains the same and is such that it allows its target to change to some degree without changing the content itself. If one accepts, first, One Space-ism or Two Space-ism with Metaphysical Plenitude and, second, that true identities are necessary, then the only way to reach the desired conclusion in Chalmers' linguistic framework is to take a descriptivist stance. The content of a belief concerning water is given by a description of the associated identifying properties and the designation of *water* is determined by that description in every possible world, that is, the interpretation function is constrained by the description. The relevant description has something to do with the observable properties of water (is clear, is liquid, falls form the sky, fills rivers, lakes and oceans, etc.) while the content of belief concerning H₂O has to do with some

other identifying properties (perhaps properties having to do with hydrogen, oxygen, molecules, atoms, etc.). Thereby the Mariinsky Theater would no longer be qualified as a designation of *water*. Instead the designations of *water* would be liquids resembling water. Consider the Twin-Earth case again: in Twin-Earth rivers and lakes are filled with XYZ. According to Jackson and Chalmers *There is water* is true and *There is H*₂O is false in Twin-Earth from the perspective of Twin-Earth and this may be employed to capture a perspectival aspect of the content of *There is water*. But a supporter of the meta-semantic interpretation of this very same 2D model would say that this does not have anything to do with a perspectival content of *There is water*. The belief that *There is water* merely has one content in Earth and another completely different content in Twin-Earth.

Which interpretation is more plausible? The 2D framework does not provide any insight concerning this matter. At the formal level we may explicate the designations of the term (or the concept) *water* in the Twin-Earth case by stating that *water* designates H₂O in Earth and XYZ in Twin-Earth. Then we may introduce the 2D tools and explicate the designations in the style favored by the two-dimensionalists and state that *water* designates H₂O in the 〈Earth, Earth〉 pair and in the 〈Twin-Earth, Earth〉 pair but XYZ in the 〈Twin-Earth, Twin-Earth〉 pair. Now we may evaluate designations in two ways, either in pairs in which Earth remains constantly as a second component, or in pairs that consist of the same world twice over. Then, in the first case, we may state that from the perspective of Earth *water* designates H₂O both in Earth and in Twin-Earth but, in the second case, water designates XYZ in Twin-Earth from the perspective of Twin-Earth.

But does this depict how the content of the belief There is water changes and how it depends on the different points of evaluation or does it depict a certain feature of the perspectival content of There is water that tracks different substances from one point of evaluation to another? This is a question that cannot be settled by using the 2D framework. It must be settled by philosophical considerations before resorting to 2D semantics. Chalmers needs to resort to descriptive beliefs and to associated identifying properties right at the outset. There is water comes with one set of associated representational properties and There is H_2O comes with some different set of properties. Hence both beliefs are purely descriptive. This contrasts with the externalist approaches inspired by Kripke (1972/1980) and Putnam (1975), which argue that the only relevant content of this sort of de re belief concerning natural kinds is the wide-content that is determined externally, not descriptively and, second, that in practice speakers and believers simply do not associate descriptions or identifying properties of the required type with natural kinds. Therefore both Jackson (2004, 263) and Chalmers (2002b, 149, 160) seek to distance themselves from descriptivism. But this puts Jackson and Chalmers in a difficult position since the 2D framework delivers the desired results only when paired with a general philosophical position reminiscent of descriptivism. Chalmers tries to settle the matter by suggesting vaguely that the meaning of a description can "approximate" the meaning of the original expression or concept

(see, e.g., Chalmers 2002b, 149, 160). It is far from clear, however, what this suggestion means exactly.

The main point of this lengthy section was that Chalmers' 2D is a complex framework that faces interpretational problems with its relation to descriptivism and the question concerning the nature of worlds. Below I sketch an outline of a Hintikka-style framework that is more straightforward in the sense that it is openly descriptivistic and invokes only metaphysically possible worlds.

4.6 EMPTY SECONDARY CONTENT

Sandgren (2018) pointed out a problem regarding the role of the secondary dimension in Chalmers' account of communication, agreement, and disagreement. As mentioned previously, Chalmers appeals to the same secondary content in these connections. The problem appears when communication, agreement and disagreement involve non-existent objects, such as Vulcan or Santa Claus. In cases such as these, the primary content does not determine any target in the actual world. Then there is either no secondary content at all, or the secondary content is captured by the empty set of worlds. In both cases the phenomena are left unexplained. In the first case Package Delivery Model #3 collapses into Package Delivery Model #2 since the slot for the secondary dimension is left empty, and there is no secondary content to appeal to.

In the second case, we would have a secondary content to appeal to, but it would be the empty set in every case when communication involves a non-existent target, that is, all the beliefs that involve non-existent targets would have the same secondary content and thereby all processes of communication that involve non-existent objects would automatically be successful according to the relevant model. For example, if I try to communicate that Vulcan is a planet and you acquire a belief that Santa Claus is coming to town, then this would be successful Chalmersian communication becausedue the secondary content of these beliefs is the same. But this is, of course, absurd.

4.7 HINTIKKA-STYLE COMMUNICATION

Here is a sketch of a Hintikka-style account that aims to explain some of the phenomena discussed. This is merely a rough proposal for a framework. I shall not define semantics in full detail, let a alone a full logic, because many intricate questions identified in the literature relating to relativization, subjectivity, reflexivity, immunity to misidentification, hidden or unarticulated constituents and where to place them (mind, language, world), and so forth, need to be studied further first (see, for instance, Recanati 2007). The sketched framework depicts epistemic possibility in terms of metaphysically possible worlds, and hence avoids commitments to more obscure epistemically possible worlds that are

metaphysically impossible. But this, of course, raises the question of what metaphysically possible worlds are. I shall not address this question. The case I try to make is that a framework built out of the elements introduced in Chapter 3 offers similar prospects in the area of perspectival communication as Chalmers' 2D. The view considered here recognizes two kinds of content, just as Chalmers' account, but it is more straightforward in its commitment to descriptivism and metaphysically possible worlds, it characterizes properly perspectival contents, it is readily implemented in Hintikka-style epistemic logic, and it does not appeal to secondary content in the way Chalmers does, namely the way criticized by Sandgren. The dimensions of content are represented here with the aid of quantifiers. I will concentrate here on the modal languages studied in the previous chapter, namely languages without individual constants, and focus on the simplest quantified formulae involving the most basic artifacts of our modeling, namely variables and predicates. Many interesting applications call for more, for instance individual constants, but here my aim is to discuss ideas and clarify the core of the framework. If that succeeds, then further embellishments may be added later and many interesting applications, for instance in the context of Fregean puzzlescenarios, may be studied in detail later.

We need semantics that incorporates Hintikka's ideas about individuation familiar from Chapter 3. In particular, we need semantics that depicts "individuation without reference and reference without individuation", as Hintikka (1969, 106) envisaged. Here standard intensional semantics for quantified modal logic will take us a long way. Such a theory assigns semantic values from a category of entities known as intensions, typically Carnapian individual concepts. The relevant intensions in Hintikka's case are, of course, world-lines. Both individuation and reference are represented by world-lines, and to achieve Hintikka's vision we must be able to employ world-lines as semantic values irrespective of whether the world-lines as functions have values themselves, that is, a world-line may not pick out anything in a given world but it may nonetheless serve as a semantic value of a variable or a predicate when evaluating beliefs and other epistemic attitudes. This allows us to model epistemic subjects that think about non-existent objects such as unicorns. ⁵⁶

World-lines, in general, are members of **F**, that is, functions of the kind characterized in Chapter 3. Intensional semantics for quantified modal logic is familiar from a number of studies in the tradition of Montague semantics as well as from more recent studies such as Aloni (2005) and Belnap & Müller (2013a). I shall not go into the details of intensional predication and basic intensional quantification. Instead, I jump right into the question of how to introduce Hintikka's two kinds of quantification.

Both types of quantification invoked by Hintikka represent intentional thoughts directed at particular objects. In order to have an object directed thought

⁵⁶ Many argue that unicorns exist as mythical or fictional objects. If so then we may alter the example and consider, for instance, my non-existent brother, concerning whom there are neither myths nor fiction, or any non-existent thing about which there seems to be true modal claims.

one has to single out an object, for instance, in one's visual field. In order to have a belief about an object, to refer to an object, to pick it out among other things, or to state something about it, one must individuate it first. The most basic kind of belief is perceptual belief involving perceptual reference, the most basic kind of reference. Direct referentialists influenced by Kripke seldom consider perceptual reference because what we perceive are properties, not objects as such, and considerations of identifying properties become relevant. According to Hintikka (1969, 151–183), perceptual reference operates via properties and an object is a property cluster from a perceptual point of view. Underneath the properties, an object as such is a construction of our cognitive abilities in a neo-Kantian fashion, as witnessed in the previous chapter. We are causally related to our environment in many ways but not directly to the objects, as a careless direct referentialist might claim, but first and foremost to the properties and only derivatively to the objects (which are constructions of our cognitive activity). After all, objects enter into causal interactions via their properties. Hence the immediate objects of perceptual belief are not objects or "propositions" but properties or property-clusters. Does this mean that there are no ordinary objects and we never have epistemic attitudes directed at them? No, it is just that our relationship to objects is more complicated and less straightforward than some currently influential approaches in philosophy are apt to recognize, and that ordinary objects themselves, or individuals, as Hintikka preferred to call them, are based on more basic primitive components such as properties and appearances.

Let us now turn to perspectival and more objective contents. Instead of Chalmers' primary and secondary content, we appeal to two different modes of individuation of objects, the perspectival and the physical/public mode. Formally, we represent these with four different quantifiers, the perspectival quantifiers Ex, Ax, and the physical/public quantifiers $\exists x$, $\forall x$ (Hintikka 1969; 1975).

One aspect that the distinction appears to reflect is the so-called duality of belief and the resulting ambiguity of belief reports. Beliefs may be assessed either by their relation to the facts (or other beliefs held by other people, other beliefs held by the same subject at different times, situations etc.) or by their "intraindividual psychological role" when the focus is instead on the internal connections between these beliefs and other states and or acts of the same subject (Recanati 2007, 113). In the first case it is the truth conditional properties of the beliefs that we are interested in and in the second case, it is the explanatory role of the belief in understanding the subject's behavior, cognition, perceptions etc.

In Hintikka's work, the intra-individual perspective was always important. There are many ways to understand the tasks of epistemic logic, perhaps as many as there are epistemic logicians, but the seminal epistemic logic of $K \mathcal{C}B$ was a project to model the information that a subject has and to model how she reasons with it from her perspective. As witnessed in Chapter 2, Hintikka's epistemic notions of defensibility, indefensibility and self-sustainability were subject-relative: what is indefensible for one subject may be defensible for another subject. Hence there was an emphasis on the intra-individual psychological role of epistemic

attitudes right at the outset. From the very beginning of *K&B*, Hintikka held that if someone believes that Dr. Jekyll is not Mr. Hyde, then there must be coherent epistemic alternatives open to this subject compatible with the belief that Dr. Jekyll is not Mr. Hyde. The facts (whether or not Dr. Jekyll is Mr. Hyde, whether or not it is possible that Dr. Jekyll is not Mr. Hyde, or whether or not Dr. Jekyll and Mr. Hyde exist at all) are irrelevant to this analysis of reasoning: if a subject believes that Mr. Hyde is a murderer and the fact of the matter is that Mr. Hyde is Dr. Jekyll, then we are not licensed to report that the subject believes that Dr. Jekyll is a murderer (Hintikka 1962, Ch. 6), not even from any objective *de re* third-person point-of-view. Even if the subject *knows who* Dr. Jekyll is and *knows who* Mr. Hyde is, we are not licensed to conclude that the subject knows that Dr. Jekyll is Mr. Hyde, apparently because the perspective from which the subject knows Dr. Jekyll might not involve him as being known as Mr. Hyde and vice versa.⁵⁷ In such a case the subject simply would not be in a position to draw the identity conclusion because of the perspectives involved.

At the level of quantifiers, this intra-individual point of view manifested in the fact that a construction beginning with $\forall x$ does not mean "for all x" and $\exists x$ does not mean "for some x" in $K\dot{c}$ ". Instead $\forall x$ means something along the lines of "for all known/believed individuals", that is, quantification concerns objects that a given subject has beliefs about, individuals that are already recognized in the belief system of that subject. A similar restriction applies to the existential quantifier (Hintikka 1962, 155).

Further, quantifying into an epistemic context, $\exists x K_a(x = Dr. Jekyll)$, does not determine a unique object in the possible worlds introduced by the modal operator K_a (contrary to many Hintikka's claims); instead the formula states that in each accessible epistemic alternative world there is some individual who is considered by the subject as being Dr. Jekyll (see the discussion in the section 2.9).

Some studies concerning the development of Hintikka's views suggest that after $K\mathcal{C}B$ Hintikka supplemented his framework by introducing world-lines, especially perspectival world-lines with novel quantifiers, perspectival quantifiers (see e.g. Tulenheimo 2009, section 4). In my view, Hintikka had been operating with perspectival quantification all along; it was the public aspect of quantification that was introduced in order to account for some of the objective properties of epistemic attitudes, and especially in order to compare the attitudes of different subjects.

Let us now assess world-lines in some detail. It is not a straightforward task. As already mentioned, Hintikka provided only general and informal explanations of the nature and the role of world-lines in quantification. As we saw in Chapter 3,

⁵⁷ As witnessed in section 2.10 a formula representing reasoning from (i) factual identity of *a* and *b*, (ii) an agent's epistemic attitude of *knowing who a is*, and (iii) *knowing who b is*, to a conclusion that (iv) the agent knows that *a* and *b* are identical, namely

 $^{[(}a = b) \land \exists x K_i(x = a) \land \exists y K_i(y = b)] \rightarrow K_i(a = b)$

is not valid in the $K \otimes B$ system. It was argued in 2.10 that this was in fact an oversight, not a deliberate attempt to take perspectives into account.

Niiniluoto (1979; 1982) provided roughly the following model-theoretic account of the modes understood as two types of quantification: O_a is a modal attitude operator relative to a subject a, while $\mathcal{O}_a(\mu)$ is the set of \mathcal{O}_a -alternatives to the actual world μ in Ω that a subject relative accessibility relation R_a provides access to.

Physical/Public Quantification:

 $\exists x O_a Fx$ is true in μ if and only if there is a physical/public world-line f on $\mathcal{O}_a(\mu)$ such that, for all $\mu^* \in \mathcal{O}_a(\mu)$, $f(\mu^*)$ satisfies Fx in μ^* .

Perspectival Quantification:

 ExO_aFx is true in μ if and only if there is a perspectival world-line f on $\mathcal{O}_a(\mu)$ relative to μ and to a such that, for all $\mu^* \in \mathcal{O}_a(\mu)$, $f(\mu^*)$ satisfies Fx in μ^* .

But what is the difference between these two types of world-lines, exactly? Traditionally it has been considered that the criteria that an individual concept operates on may be specified and spelled out by a definite description such as *The President of the United States*. Perhaps the difference between the two is just that the perspectival world-lines are indexicalized to an agent by the irreducible presence of "T" and "now" and such in their specifications – *the stuff I happen to be looking at right now, the thing right in front of me that seems to me to be blue* – while the public world-lines operate on more objective criteria, such as *the stuff that is* H_2O . However, Hintikka (1998, 205) held that world-lines may represent anything that may plausibly be taken as a mode or method of individuation, and some of these modes may not involve descriptive elements but rather "an agent's first-hand cognitive relations" to objects. If this is correct, then perhaps some appropriate causal relations or immediate perceptions may serve as examples.

There are many ways to accommodate the two sets of modes of individuation in one's formal semantics. One way which makes world-lines count, and which is also in line with Hintikka's neo-Kantianism, is to simply consider world-lines as individuals, that is, to consider the domain D as consisting solely of both types of world-lines. In section 4.5 above, we considered D as consisting not of world-lines but of ordinary objects such as people, planets, and household items, and we considered each sub-domain D_{μ} as consisting similarly of people, planets, and household items existing in each relevant world. How about in the case of world-lines? If D consists of world-lines, then no D_{μ} may consist of world-lines, since world-lines do not exist in any particular world. This seems to force on us a distinction between individuals (world-lines) and world-bound objects existing in possible worlds.

The way I understand the situation is that, in this connection, we should appeal to the neo-Kantian background view and consider world-bound objects as more basic primitive components such as properties and appearances that the individuals proper are based on. Hence a world-bound object existing in some possible world is not an individual in a full sense but a cluster of properties and appearances, a perceivable or a thinkable target that may possess some further features. These world-bound objects are raw material of individuals that emerge when a world-line representing a mode of individuation (a cognitive capacity of a thinking subject) connects the world-bound objects into a unity, that is, when a world-line individuates an individual. If so, then the traditional concepts of rigidity and trans-world identity are ruled out right at the outset. The sameness and identity conditions required for these notions apply to individuals proper. They do not apply to world-bound objects. A world-bound object in μ is just a bundle of properties and appearances, and any world-bound object in μ^* is just another bundle. Individuals may emerge only when the cognitive activity of a thinking subject constructs an individual from the bundles.

One way to proceed from here is to consider D as the domain of quantification consisting of world-lines, D^* as the domain of world-bound objects, and each D_u^* as a sub-set of D^* and as a sub-domain consisting of world-bound objects. Then a world-line is essentially a function from the set of worlds W into D^* . This seems to be the background idea in Niiniluoto's semantics. Reconsider Niiniluoto's clauses above. Both end with the requirement that $f(\mu^*)$ satisfies F_X in μ^* . I take this to mean that $f(\mu^*)$ is the value of the world-line f in the world μ^* and that value, a world-bound object existing in μ^* , must satisfy Fx, that is, that the worldbound object must be assigned in the extension of F in μ^* . So what is going on here is that our variables range over elements of D and our predicates apply to completely different entities, namely to elements of D*. I suppose that before the evaluation the extension of F has been already defined in μ^* by the means of a sub-domain of world-bound objects in μ^* and the interpretation, as is customary. Assumedly, then, some world-bound objects in μ^* are in the extension of F and some objects are not. But considering Hintikka's presuppositions this seems unintuitive. Our neo-Kantian individuals are world-lines, and surely we should be concerned with their properties first and foremost. Accordingly, we should have predicates that apply to individuals, that is, world-lines, and our general formulae should be true of world-lines rather than world-bound objects.

I suggest a more intuitive (and simpler) way to proceed. My proposal is to simplify Niiniluoto's semantics and to assume only one constant domain D consisting of world-lines that variables range over, and let predicates also apply to its elements. The interpretation function I operates in the case of predicates in a standard manner, the only exception being that elements of D assigned in extensions of predicates in each world are not world-bound objects but world-lines. This indeed stems from the idea that the modal formulae should be evaluated relative to modes of individuation. Hence, in general, it is simpler and more consistent with the presuppositions concerning world-lines to let D consist of world-lines and let predicates apply to its elements while defining a variable assignment g as a function mapping variables into D. As already mentioned, for simplicity the only singular terms considered here are variables. The designation

 $[t]_{M,\mu,g}$ of a term t in a model M at a world μ with respect to g is a world-line given by

$$[t]_{M, \mu, \rho} = g(t)(\mu)$$

while the interpretation of predicates is such that for each *n*-placed predicate symbol P_i^n and a world μ ,

$$I(P_i^n, \mu)$$

is a set of *n*-tuples of world-lines. Now, intuitively, the extensions of predicates must be allowed to vary from world to world. As Hughes & Cresswell (1996, 344ff) argue, such models with intensional predicates and intensional quantification may come with metaphysical repercussions but they do not affect the logic itself: simple models of the form $\langle W, R_i, D, I \rangle$ do not have any constraints on what the elements of D may or may not be. Even if the elements of D are intensional entities, the resulting semantics and the definition of valuation remain standard. The only exceptional aspect here is "metaphysical": the elements of D considered here are world-lines.⁵⁸ The assignment is subject relative and

⁵⁸ So the frame of the envisaged semantics would simply be $\langle W, R_i, D \rangle$. More complicated semantics involving both world-bound objects and individuals (world-lines) would explicitly recognize the respective distinction by involving domains for world-bound objects D^* and for the two sets of world-lines \mathcal{I}_1 and \mathcal{I}_2 . For instance, a frame might be $F = \langle W, R_i, D^*, \mathcal{I}_1, \mathcal{I}_2, \mathcal{Q} \rangle$ where W is a set of worlds, R_i is an accessibility relation, D^* is a domain of world-bound objects, and \mathcal{I}_1 and \mathcal{I}_2 are sets of public and perspectival world-lines respectively. \mathcal{Q} would be a domain function assigning to each possible world w its own domain \mathcal{Q}_w of world-bound objects. For every $i \in \mathcal{I}_1$, $i(w) \in \mathcal{Q}_w$ if i(w) is defined in w, and in Hintikka's terms q is a "manifestation" of i in w if q is such that i(w) = q. The same applies for every $i \in \mathcal{I}_2$. Since the resulting framework would immediately invoke constants and variables for both world-bound objects and world-lines, an assignment g would be defined as a function mapping "object variables" into \mathcal{D}^* and "world-line variables" into \mathcal{I}_1 and \mathcal{I}_2 ; and an interpretation I would give the value of constants and predicates. Then we would start from the bottom, that is, from constants and variables designating world-bound objects, and build our way gradually towards world-lines and their respective variables. The definition of designation of singular "object terms" would be:

$$[t]_{M, w, g} = I(t)(w) \in Q_w$$
 where t is a "object constant" $[t]_{M, w, g} = g(t)(w) \in Q_w$ where t is a "object variable"

The interpretation of these constants and variables is not rigid, in accordance to our neo-Kantian orientation. The interpretation of a predicate would be

$$P_{w''} \subseteq Q_{w''}$$
 (P is an n-placed predicate)

Semantics proper would be defined with respect to a model M and an assignment g (for simplicity, I ignore the question of existence):

$$\begin{split} M, \, w, \, g &\models \, P(t_1, \dots, t_n) \qquad &\text{iff} \qquad & \langle [t_1]_{M, \, w, \, g}, \, \dots, \, [t_n]_{M, \, w, \, g} \rangle \in \, P_{w''} \\ M, \, w, \, g &\models \, t = t' \qquad &\text{iff} \qquad & [t]_{M, \, w, \, g}, \, [t']_{M, \, w, \, g} \in \, Q_w \text{ and } \, [t]_{M, \, w, \, g} = [t']_{M, \, w, \, g}, \end{split}$$

Logical constants would follow in a standard manner, but when we would reach our desired "world-line variables" and quantification over world-lines and state that

$$M, w, g \models (\exists x) \varphi$$
 iff there is at least one $i \in \mathcal{I}_1 \dots$

we would notice that our variables range over elements of \mathcal{I}_1 and \mathcal{I}_2 but what we have said so far is that our predicates apply to world-bound objects, i.e. elements of D^* . This would be coherent but somewhat unintuitive for the reasons stated previously, and it would only complicate things. As Hughes & Cresswell (1996, 344ff) point out, if there is a framework that *quantifies* over intensional entities, such as world-lines in our case, then there seems to be no reason why predicates should not apply to world-lines. Indeed, we may drop all reference to D^* and assign to every n-ary predicate as its semantic value a subset of pairs of an n-tuple

situation relative. We are interested here in epistemic attitudes. An epistemic attitude requires an epistemic subject. Assignments reflect some of the subject's conceptions regarding the targets of her epistemic attitude in a certain situation, namely those conceptions that relate to the individuation of targets. Some of the conceptions are personal, private, or subjective, while some are shared, public, or objective. Therefore a typical predicate comes with both perspectival and public world-lines assigned as its semantic value (or assigned to its extension, but this is not perhaps the best choice of terminology in the present intensionally oriented setting).

The two modes of quantification provide a way to talk about beliefs that involve differently individuated targets. In a sense, perspectival quantification carries out the same task as Chalmers' primary dimension, but here the general idea is explicit: what matters for the perspectival content is the how the belief-targets are individuated.

Let us consider an example familiar from the previous sections, namely the belief that there is water. This belief may be represented in several ways. Since we have been assessing a language without individual constants, we cannot invoke kind names in this connection. Nevertheless, if a is an epistemic subject and W is shorthand for *water* then one way to state that a believes that there is water would be

$$\exists x B_a W x$$

The truth of this formula, according to our intensional semantics, requires that there is a public world-line such that it is assigned in the extension W in all relevant possible worlds introduced by the belief operator. The intended intuitive meaning of this formula is that there is some thing that is individuated in a public mode, and that a believes that this thing is, or has the property of being, water. To assess the formula more closely, we need to assume a context. Many relevant aspects are left unarticulated here and are therefore somewhat vague. A positive aspect is, however, that the formula is very simple and manageable compared to formalizations such as those in Holliday & Perry (2014) that seek to explicate all epistemically and alethically relevant matters at the level of formal language. The attitude reports of Holliday & Perry (2014) come with specific semantics and thereby possess a very precise meaning ,but the downside of their approach is that their formulae become highly complex.

In any case, at Hintikka's chosen level of abstraction the use of the public quantifier indicates that the interest here lies in the objective aspects of the belief,

of world-lines and a world. This simplifies things considerably, and our logical and metaphysical views concerning what may be predicated to what would not constrain our framework, that is, we do not have to worry about invoking illogical or metaphysically impossible worlds. Most importantly, since we are not here interested in world-bound objects at all (they are not full individuals, really) we may drop D^* from the frame, and define D as consisting of elements of \mathcal{I}_1 and \mathcal{I}_2 as I suggested above.

⁵⁹ Hintikka himself would have preferred a formula such as $\exists x B_a(x = water)$ but here water would be an individual constant.

such as whether or not the belief is true in the situation in which it is evaluated, and how, for instance, the belief relates to other beliefs held by other subjects or to other beliefs held by the same subjects at different times, and so forth. Given that water is H₂O, and that it is common knowledge that water is H₂O, it is natural to suppose that the mode of individuation involved has something to do with the target being H₂O (and perhaps some other factors) and not some other liquid or similar substance. Perhaps the context in which this epistemic attitude occurs is "scientific". Hence the relevant world-line (function) operates on the condition that, whatever the function picks out, it must have the property of being H₂O, and the formula requires this in every world introduced by the operator that world-line is to be found and it is assigned in the extension of W. In a sense, this mode of individuation represented by a function is not a part of our semantics proper, rather it is the semantic value of a variable and thus a sort of "metaphysical item" to which our predicates also apply. In any case, the division in modal space that corresponds to this belief should be such that the worlds that do not contain H₂O are excluded. Thereby an objective content of that belief is represented, and we may evaluate it in relation to a given world and in relation to beliefs employing a similar mode of individuation held by other subjects. In a world with no H₂O, this belief should be false together with other beliefs that employ a similar mode of individuation.

What if the reported epistemic attitude occurred in the 18th century? One option is to think that then the public individuation in question has something to do with the most commonly recognized properties of water, such as: is transparent, is potable, fills the lakes and rivers and so forth. In this case the individuation is less precise and the world-set in which this belief is true includes worlds with no H₂O. Existence of any transparent, potable, etc. substance suffices. Another way to represent the belief that there is water is

$$E y B_a W y$$

The truth of this formula requires that there is a perspectival world-line such that it is assigned in the extension W in all relevant possible worlds introduced by the belief operator. The intuitive meaning is that there is some thing that is individuated from the perspective of the subject a, and that a believes that this thing is water. The use of a perspectival quantifier suggests several things: first, the interest here lies in the intra-individual psychological role of the belief. The truth of the belief is irrelevant. Second, a's mode of individuation is not necessarily shared by other subjects and therefore it is natural to suppose that it has something to with some superficial properties of water and how they appear to the subject in a given situation. In other words, the world-line (function) operates on criteria that have to do with some subjective appearances of water relative to a in some situation. The function picks out things with similar appearances in the range of relevant possible worlds. If the appearances in question include being clear and transparent, then the function may from the public point of view pick out, for

instance, H_2O in one world and pure ethanol in another world. Hence the division in modal space that corresponds to this belief is such that the worlds that do not contain H_2O but similar substances are included. The belief may be true in world with no H_2O because it is evaluated relative to the mode of individuation, and the perspectival mode may not require that the target must be H_2O . The appearances may be peculiar, esoteric, dreamlike, or even hallucinatory. Thereby a subjective or a perspectival content of the belief is represented. The represented perspectival content may or may not be limitedly accessible in a Fregean fashion, that is, it may be such that only a may grasp it.

Assumedly a subject may entertain a public belief without a perspective and vice versa. The perspectival quantification captures the subjective intra-individual aspect of the relevant belief, while the public quantification captures the more objective, shareable aspect of the belief. Since we have identified a perspectival and a public dimension of the belief in question, we are in a position to give the first tentative solution to the problem of communication at hand: we proceed as Chalmers and appeal to the public aspect of belief in explaining communication involving perspectival contents. Instead of the pair of $\langle primary content \rangle$, secondary content we appeal to the pair $\langle perspectival content, public content \rangle$ and employ model #3 as Chalmers does. A nice feature of quantifiers is that they may be iterated. If an application calls for two or more perspectival or public contents or any combination thereof, we are free to stack quantifiers as we please. One way to express the outcome of successful communication with a single formula involving the speaker S and the hearer H as epistemic subjects and There is water as the belief that has been communicated is

$$\exists x E y (B_s(Wx \land Wy) \land B_H Wx)$$

The truth of the formula requires that there is a public world-line and that there is a perspectival world-line such that the public world-line is assigned in the extension of W in the possible worlds introduced by the both operators B_s and B_{tb} and that the perspectival world-line is assigned in the extension of W in the worlds introduced by B_s . The intended intuitive meaning is that there is some thing that is individuated in a public mode shared by the speaker and the hearer and that they both believe that this thing is water. In addition, there is a perspectivally individuated thing of which the speaker believes that it is water (and as it happens, the speaker believes that the hearer believes the public belief in question, which is a plausible outcome since that was the belief or content just communicated). The situation may also involve a perspectival belief of the hearer but let us ignore that to keep things manageable. In a typical case, we may assume that the public mode has got something to do with water being H_2O , and that this content, captured by the set worlds with H_2O in all of them, is the shared public content, while the perspectival mode has got something to with the superficial properties of water. In

addition to communication, this approach may be applied directly to agreement and disagreement.

But what if the situation is not typical along the lines above? What if the appearances relevant to the perspectival mode are so peculiar, esoteric, dreamlike, or hallucinatory that they do not have anything to do with the public criteria? The framework as it is laid out so far does not constrict the perspectival individuation mode in any way. What if the perspective of a given subject is so outlandish that it may be represented by a world-line that picks out, for instance, the Mariinsky Theater, Hermitage, and the Tretyakov Gallery, and then this world-line is assigned in the extension of W? Then, apparently, $\exists x B_a W x$ would be true at worlds in which there is H_2O , and EyB_aWy would be true at worlds in which there are Russian cultural institutions. In that case it seems that there are two completely different contents rather than a public content and a perspective on it. We might appeal to the initial intentional state of the subject and insist that the assignment reflects the speaker's understanding of her environment in a contorted epistemic situation, and our model depicts an intentional state that comes with a highly unusual perspective. However, as in the conflict of the Chalmersian and the meta-semantic interpretations of 2D described in section 4.5 above, it may be claimed that rather than capturing a perspective the formula depicts the fact that the concept water is ambiguous and/or vague in the speaker's belief system, or it would depict the fact that the predicate water in the mouth of the speaker is ambiguous and/or vague.

Our explanation would still work because it ascribes the public belief to the speaker and to the hearer and that is what gets communicated. However, it may be argued that the perspectival belief would not represent a perspective but another belief that is not relevant to the purposes at hand. If we are interested in capturing a perspective that is more intimately linked to public criteria then one thing we may try is to invoke formulae involving identity. Niiniluoto (1982) employed in perceptual contexts formulae such as

$$\exists x E y S_a(x = y)$$

where S_a is perceptual operator "a sees". Such formulae are also interesting for our purposes. But there are two problems: Niiniluoto's clauses do not tell us exactly how we should interpret formulae such as $\exists x E y O_a(x=y)$. After we are done with the first quantifier, we confront another one and the clauses above do not help us in this situation. Another problem is that the truth of $\exists x E y O_a(x=y)$ requires that there is a public world-line and that there is a perspectival world-line such that in the range of possible worlds introduced by the operator O_a these world-lines are identical. What exactly does it mean for a public and a perspectival world-line to be identical?

In the case of the first problem we may simply introduce a new hybrid clause in the spirit of Niiniluoto:

Hybrid Quantification:

 $\exists x E y \, O_a(x = y)$ is true in μ if and only if there is a public world-line f_1 on $\mathcal{O}_a(\mu)$, and there is a perspectival world-line f_2 on $\mathcal{O}_a(\mu)$ relative to μ and to a, such that, for all $\mu^* \in \mathcal{O}_a(\mu)$, f_1 and f_2 satisfy (x = y) in μ^* .

Now we at least have instructions to read $\exists x E y O_a(x=y)$. 60 The second question was: what is it for world-lines to satisfy (x = y) in some range of possible worlds? A trivial answer is that f_1 and f_2 satisfy (x = y) in a given world μ just in case the interpretation function assigns f_1 and f_2 in the extension of the intensional identity predicate in μ . While this is perfectly sufficient for semantic purposes, one might want a more substantial answer. A more substantial answer would probably appeal to the fact that world-lines are functions, and functions may be understood as extensional in the sense that two functions are identical in μ just in case they deliver the same value in μ . This allows that functions may operate on different criteria but nevertheless be identical, and that is exactly what we want. A function representing public individuation criteria and a function representing perspectival criteria may be identical in a certain range of worlds if they deliver the same values in those worlds. Perhaps it would be better to say that the functions and thereby the world-lines coincide, rather than to say that they are identical. But making this explicit would again require a distinction between the elements of D, that is worldlines, and the values that they take in each possible world, that is, further domains of world-bound objects existing in each possible world. This would force us to abandon the chosen level of abstraction and again lead to intricate questions concerning the neo-Kantian nature of these objects and the exact form of our semantics. I do not assess these questions here in detail. I think the identity of world-lines is sufficiently clear and we have a pretty good intuitive idea of what it means. The world-lines are taken as primitive semantic values of variables and here I have tried to elucidate their nature to a sufficient degree. The exact questions regarding their properties and identity conditions are metaphysical, and I cannot go into these rich details here.⁶¹

⁶⁰ Niiniluoto does not define semantics in full detail. As pointed out earlier, we need, in addition to this clause, at least a definition of assignment, and since the evaluation of such formulae starts from the inside, that is from x = y, we would also need a notation that indicates whether the variables are interpreted physically/publicly or perspectivally. Then an assignment would assign corresponding world-lines to free variables accordingly. But I assume that the general idea is sufficiently clear without departing from Hintikka's original notation.

One way to address some of the metaphysical questions would be, perhaps, to further illustrate the nature of elements of D by defining a mode of individuation, or a world-line, in the context of counterpart semantics compatible with Hintikka's neo-Kantianism. Let us introduce a set of objects D_0 , i.e., a set of world-bound objects that is not a part of our semantics proper but serves to illustrate the nature of the semantic values that we assign to variables. In its most general form a mode of individuation is a relation $\mathscr M$ between ordered pairs $\langle w, d \rangle$ of worlds in W and individuals in D_0 . To state that

 $[\]langle w, d \rangle \mathcal{R} \langle w', d' \rangle$

is to state that d' is a counterpart of d in w'. Further constraints may be set for that relation: for example, we may require that the relation is functional, i.e. that one object in one world has at most one counterpart in

Hence, yet another idea on how to use world-lines in capturing perspectives is to stack the two quantifiers, include the identity symbol, and utilize formulae obtainable from the schema

$$\exists x E y B_a((x = y) \land \varphi))$$

where B_a is a belief operator and φ is any formula representing what is believed and containing bindable variables. The truth of the part before the conjunction requires that there is a public world-line (a function operating on some objective criteria) and that there is a perspectival world-line (a function operating on some subjective criteria relative to a) such that these world-lines are identical in the range of possible worlds compatible with what a believes (i.e. the functions are identical, i.e. the functions have the same values). The intuitive meaning of such formulae is that the targets of the relevant belief are individuated in both public and perspectival mode and that these targets thereby doubly individuated must fulfill both perspectival and public individuation criteria in the range of possible worlds introduced by the operator. The perspectival quantification captures the subjective intra-individual aspect of the relevant belief while the public quantification captures the more objective, shareable aspect of the belief.

In our example case, the use of such formulae comes into question when the epistemic subject is well informed and the belief is formed in and ideal or at least in more or less epistemically charitable situation. Hence yet another way to represent the belief that *There is water* relative to a subject *a* would be

$$\exists x \to y B_{\alpha}((x = y) \land Wy))$$

The truth of the formula requires that there is a public world-line and that there is a perspectival world-line relative to *a* such that, in the range of possible worlds compatible to what *a* believes, the public world-line and the perspectival world-line are identical, and in that range of worlds the perspectival world-line is assigned in the extension of *W*. The intuitive meaning of such formulae is that the targets of the relevant belief are individuated in both public and perspectival mode and that these targets thereby doubly individuated are the same in the range of possible worlds introduced by the operator. The perspectival quantification captures a subjective intra-individual aspect of the relevant belief, while the public

another world. In such a case we would have a typical basic world-line. However, Hintikka insists in many connections that the world-lines may "split", that is, one mode of individuation may recognize more than one world-bound object in a world. If Hintikka is right, the functionality restriction is too strong. Then we could say that a minimum formal requirement for perspectival world-line is just that it is a relation of the above kind. Then a perspectival world-line similar to an individual concept would come with the functionality constraint. Candidates for the minimum formal requirements for a public world-line would be, first, that it is a relation of the above kind, second, it is a functional relation, and third, it is not a partial function but total in the sense that, if an object has an counterpart in one world, then it has a counterpart in every relevant world. I do not suggest that all there this is to a mode of individuation is a constrained relation between W and D_0 . In addition to the minimum formal requirements, we may, of course, want to set many other requirements.

quantification captures a more objective, shareable aspect of the belief. Then we proceed as Chalmers and appeal to the public aspect of belief in explaining communication, agreement, and disagreement.

Thereby a proper perspectival content relating to water and nothing else has been represented. If limited accessibility applies, then this perspectival aspect of content may be seen as such that only *a* can grasp it. Further, we have stated only that modes of individuation exist. The individuating properties by virtue of which the public individuation operates do not have to be instantiated in the actual world. Therefore beliefs such as *There are Unicorns* may be treated in this approach.

Next, I shall consider some immediate objections to this view.

1) The account of quantification is non-standard and creates unnecessary complications.

Quantification in the contexts of epistemic modality is a complex and subtle issue in any case (see e.g. Holliday & Perry 2014). Most of the available accounts are intensional because many philosophers want at least some genuine terms to be, first, non-rigid, and second, not synonymous with Russellian descriptions. If so, then neither the classical nor the free-logic quantifier rules are acceptable. A straightforward solution is to generalize the treatment of quantifiers as we did above and take the domain of quantification as consisting of intensions such as world-lines rather than world-bound objects (see e.g. Bressan 1973; Aloni 2005; Belnap & Müller 2013a, 2013b).

- 2) Quantifying-in creates existential commitments in the cases of Vulcan and Santa Claus. True. But the existential commitment is to the existence of a world-line, that is, to a mode of individuation, not to Vulcan or Santa Claus. Of course, the view is committed to the claim that we may possess a mode to individuate Vulcan, that is, the mode may exist without Vulcan existing.
 - 3) Without a secondary dimension or something similar to the neo-Russellian content, one is floating free in epistemic space unanchored to reality. One cannot accommodate the externalism of Kripke and Putnam, and one may have only purely descriptive thoughts concerning the world.

Sandgren's (2018) point establishes that we cannot appeal to a secondary content in the cases of Vulcan and Santa Claus. Communication and other phenomena in these cases must therefore be explained by resorting to other resources. Whatever the explanation might be, it is clear that these resources that we appeal to in these cases will also be available in the cases that involve existent objects. In other words, secondary content may not be necessary for explaining communication etc. There is a way, however, to approximate secondary content in other contexts with resources at hand. We need natural kind predicates with externalist satisfaction conditions. The approach at hand does not exclude them in any sense. The idea was spelled out in section 3.4: public world-lines may be tied to externalist satisfaction conditions of some theoretical term such as H_2O .

4) The distinction between the perspectival and the public mode is fuzzy. They are both perspectival, the other is only said to be less perspectival.

True. They may both belong to what Chalmers calls the primary dimension. But, as stated above, in explaining communication we must also appeal to resources outside the secondary dimension. Our distinction suggests some resources.

5) One cannot accommodate rigidity and necessary identities.

Alethic modality is not our topic here, but we may ask what kind of logic of identity of objects is available to us in the contexts at hand. Kripke models with rigid terms are out of reach from our viewpoint simply because we deny the existence of the primitive objects that are required for such models. The logic of identity ,combined with alethic modality in terms of world-lines and objects relative to the modes of individuation, would, if coherent, be very different from a Kripkean logic.

4.8 METAPHYSICALLY OR EPISTEMICALLY POSSIBLE WORLDS?

In the previous sections I further studied Hintikka's views on quantification in epistemic contexts by assessing problems relating to perspectival content. I surveyed some proposals to solve them and focused especially on Chalmers' two dimensions of content. I agreed with Chalmers that it is a promising idea to acknowledge more than one kind of content, and I showed that Hintikka's distinction of two modes of individuation may be deployed to similar theoretical purposes.

I also presented remarks concerning architectures of modal space and we witnessed some difficulties involved in "Two-Spacism", that is, in the recognition of epistemically possible worlds in addition to metaphysically possible worlds. The Hintikka-style view discussed above is an openly descriptivistic view, and hence it does not require epistemic possibilities that are not metaphysical possibilities. If a subject believes that there is water and that there is no H₂O, we do not need to invoke metaphysically impossible but yet epistemically possible worlds in which there is water but no H₂O. The worlds required by our analysis are simply metaphysically possible worlds in which we find a liquid closely resembling water which is not H₂O.

In general, the present view depicts epistemic possibilities by resorting to metaphysically possible worlds. Hence the view is well positioned to defend the modal picture of information, inquiry, and content mentioned in the Introduction. The difficult question concerning the exact nature of metaphysically possible worlds was not discussed.

5 CONCLUSION

I have surveyed Hintikka's logico-semantical works from a historical point of view as well as from a more systematic point of view in relation to some current topics with an emphasis on certain special problems related to the general notions of epistemic space, individual, and possible world. The main conclusions of the historical part were that the tool Hintikka deployed to construct epistemic spaces, namely his model set semantics, is not Kripke semantics, and that Hintikka's early modal logic has certain shortcomings. I argued in Chapter 2 that Hintikka's model set framework was built on syntactic foundations and it simply lacks some desirable semantic features. Section 2.10 showed that attempts to improve Hintikka's framework by enriching it with sufficient semantic details, as suggested in Hintikka's informal remarks in $K\mathfrak{C} B$, lead to serious new problems.

Hintikka's individuals, and especially how the notion of rigidity functions in different contexts in relation to these individuals, was the topic of Chapter 3. I argued that many of Hintikka's ambivalent remarks concerning rigidity become more comprehensible if, first, three different variants of rigidity are acknowledged and, second, Hintikka's largely implicit doctrine of semantic neo-Kantianism is made more explicit, for instance, by specifying the foundational presuppositions of Hintikka's theory of quantification.

Chapter 4 developed Hintikka's views on individuation and quantification by assessing the notion of perspectival content in the context of a more contemporary application, namely semantics involving cognitive significance and dimensions of content. In addition to a novel interpretation of the framework the chapter also suggested that Hintikka's work in the 70s anticipated the current two-dimensional approaches in modal semantics.

REFERENCES

- Acta Philosophica Fennica 16, Proceedings of a Colloquium on Modal and Many valued Logics Helsinki 23-26 August 1962, Helsinki: Philosophical Society of Finland.
- Aloni M 2005 Individual Concepts in Modal Predicate Logic, *Journal of Philosophical Logic* 34/1, 1–64.
- Auxier R E & Hahn L E (eds.) 2006 *The Philosophy of Jaakko Hintikka*, Chicago: Open Court.
- Bach K 1987 Thought and Reference, Oxford: Clarendon.
- Bach K 2010 Getting a Thing into a Thought, *New Essays on Singular Thought* ed. by Jeshion, New York: Oxford University Press.
- Barcan Marcus R 1981 A Proposed Solution to a Puzzle about Belief, *Midwest Studies in Philosophy* 6/1, 501–510.
- Barcan Marcus R 1983 Rationality and Believing the Impossible, *Journal of Philosophy* 80/6, 321–338.
- Barnes K 1976 Proper Names, Possible Worlds, and Propositional Attitudes, *Philosophia* 6/1, 29 38.
- Barwise J & Perry J 1983 Situations and Attitudes, Cambridge, Mass., MIT Press.
- Belnap N & Müller T 2013a CIFOL: A Case Intensional First Order Logic (I): Toward a Logic of Sorts, *Journal of Philosophical Logic*, doi: 10.1007/s10992-012-9267-x.
- Belnap N & Müller T 2013b BH-CIFOL: A Case Intensional First Order Logic (II): Branching Histories, *Journal of Philosophical Logic*, doi: 10.1007/s10992-013-9292-4.
- Berger A (ed.) 2011 Saul Kripke, Cambridge, Mass.: Cambridge University Press.
- Beth E W 1955 Semantic Entailment and Formal Derivability, Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen, Afd. Letterkunde, Nieuwe Reeks 18/13, 309–342.
- Blackburn P, van Benthem J & Wolter F (eds.) 2007 Handbook of Modal Logic, Amsterdam: Elsevier.
- Braun D 1998 Understanding Belief Reports, *Philosophical Review* 107/4, 555–595.
- Braun D 2002 Cognitive Significance, Attitude Ascriptions, and Ways of Believing Propositions, *Philosophical Studies* 108/1, 65-81.
- Bressan A 1973 A General Interpreted Modal Calculus, New Haven: Yale University Press.
- Burge T 2009 Five Thesis on *De Re* States and Attitudes, *The Philosophy of David Kaplan*, ed. by Almog & Leonardi, New York: Oxford University Press.
- Bystrov P & Sadovsky V (eds.) 1996 *Philosophical Logic and Logic of Philosophy*, Dordrecht: Kluwer.

- Caplan B 2005 Against Widescopism, Philosophical Studies 125, 167-190.
- Cappelen H & Dever J 2013 The Inessential Indexical, Oxford: Oxford University Press.
- Carnap R 1947 Meaning and Necessity, Chicago: University of Chicago Press.
- Castañeda H-N 1964 Jaakko Hintikka. Knowledge and Belief. An Introduction to the Logic of the Two Notions, *Journal of Symbolic Logic*, 29/3, 132–134. [A review of Hintikka (1962)].
- Chalmers D 1996 *The Conscious Mind In Search of a Fundamental Theory*, Oxford: Oxford University Press.
- Chalmers D 2002a The Components of Content, *Philosophy of Mind: Classical and Contemporary Readings*, ed. by D. Chalmers, Oxford: Oxford University Press.
- Chalmers D 2002b On Sense and Intension, *Philosophical Perspectives* 16, 135–182.
- Chalmers D 2004 Epistemic Two-Dimensional Semantics, *Philosophical Studies*, 118, 153–226.
- Chalmers D 2006a The Foundations of Two-Dimensional Semantics, in Garcia-Carpintero & Macia (2006), 55–140.
- Chalmers D 2006b Two-Dimensional Semantics, *The Oxford Handbook of Philosophy of Language*, ed. by E. Lepore & B. C. Smith, Oxford: Oxford University Press.
- Chalmers D 2009 The Two-Dimensional Argument Against Materialism, Oxford Handbook of Philosophy of Mind ed. by McLaughlin, Oxford: Oxford University Press.
- Chalmers D 2010 The Character of Consciousness, Oxford: Oxford University Press
- Chalmers D 2011a Propositions and Attitude Ascriptions: A Fregean Account, *Noûs* 45/4, 595–639.
- Chalmers D 2011b The Nature of Epistemic Space, in Egan & Weatherson (2011).
- Chalmers D & Jackson F 2001 Conceptual Analysis and Reductive Explanation, *Philosophical Review* 110, 315–61.
- Chisholm R M 1963 The Logic of Knowing, *Journal of Philosophy*, 60/25, 773–795. [A review of Hintikka (1962)].
- Chisholm R M 1976 Person and Object: A Metaphysical Study, La Salle, IL.: Open Court.
- Chisholm R M 1984 The Primacy of the Intentional, Synthése 61, 89–110.
- Copeland B J 2002 The Genesis of Possible World Semantics, *Journal of Philosophical Logic* 31, 99–137.
- Copeland B J 2006 Meredith, Prior, and the History of Possible World Semantics, *Synthése* 150/3, 373 397.
- Cresswell M J 1978 Semantic Competence, *Meaning and Translation* ed. by Guenthner F & Guenthner-Reutter M, London: Duckworth.

- Crimmins M & Perry J 1989 The Prince and the Phone Booth: Reporting Puzzling Beliefs, *Journal of Philosophy* 86/12, 685-711.
- Davies M 2004 Reference, Contingency, and the Two-dimensional Framework, *Philosophical Studies* 118, 83–131.
- Davies M & Stoljar D (eds.) 2004 The Two-Dimensional Framework and its Applications: Metaphysics, Language, and Mind, a special issue of Philosophical Studies 118/1-2.
- Dummett M 2004 Frege: Philosophy of Language, London: Duckworth.
- Egan A & Weatherson B (eds.) 2011 *Epistemic Modality*, Oxford: Oxford University Press.
- Egan A 2006 Secondary Qualities and Self-Location, *Philosophy and Phenomenological Research* 72, 97–119.
- Fagin R et al 1995 Reasoning about Knowledge, Cambridge, Mass.: The MIT Press
- Føllesdal D 1969 Quine on Modality, Words and Objections, ed. by Davidson & Hintikka, Dordrecht, Reidel.
- Føllesdal D 1994 Stig Kanger in memoriam, Logic, Methodology and Philosophy of Science IX, ed. by D. Prawitz, B. Skyrms & D. Westerstahl, Amsterdam: Elsevier.
- Føllesdal D 2016 On the Genesis of Possible World Semantics, a talk delivered at Jaakko Hintikka Memorial Conference: Legacy in Game-Theoretical Semantics and Epistemic Logic, 8. 9. September 2016, University of Helsinki.
- Frege G 1892 Über Sinn und Bedeutung, Zeitschrift für Philosophie und philosophische Kritik 100, 25–50, as reprinted in Frege 2002.
- Frege G 1918/1956 The Thought: A Logical Inquiry, *Mind*, New Series 65 (259), 289–311.
- Frege G 2002 Funktion Begriff Bedeutung, ed. by M. Textor, Göttingen: Vandenhoeck & Ruprecht.
- Gabbay D M & Woods J (eds.) 2006 Logic and the Modalities in the Twentieth Century, *Handbook of the History of Logic 7*, Amsterdam: Elsevier.
- Garcia-Carpintero M & Macia J (eds.) 2006 Two-Dimensional Semantics: Foundations and Applications, Oxford:Oxford University Press.
- García-Carpintero M & Torre S (eds.) 2016 About Oneself De Se Thought and Communication, Oxford: Oxford University Press.
- Gibbard A 2012 Meaning and Normativity, Oxford: Oxford University Press.
- Goldblatt R 2006 Mathematical Modal Logic: a View of its Evolution, in Gabbay & Woods (2006), 1–98.
- Grice H P 1969 Vacuous Names, Words and Objections, ed. by Davidson & Hintikka, Dordrecht: Reidel.
- Halonen I 2009 von Wrightin ja Hintikan rooleista mahdollisten maailmojen semantiikan syntyvaiheissa, *Ajattelun välineet ja maailmat*, ed. by J. Manninen & R. Vilkko, Helsinki: Gaudeamus.
- Hanks P 2009 On Recent Work on Propositions, *Philosophy Compass* 4/3, 469 486.

- Hegarty, M 2016 Modality and Propositional Attitudes, Cambridge: Cambridge University Press.
- Hilpinen R 2006 Jaakko Hintikka on Epistemic Logic and Epistemology, in Auxier & Hahn (2006).
- Hintikka J 1955 Form and Content in Quantification Theory, *Acta Philosophica Fennica* 8, 5–55.
- Hintikka J 1957a Modality as Referential Multiplicity, Ajatus XX, 49-64.
- Hintikka J 1957b Quantifiers in Deontic Logic, Societas Scientiarum Fennica, Commentationes Humanarum Litterarum 23/4, 2–23.
- Hintikka J 1961 Modality and Quantification, Theoria 27, 119-128.
- Hintikka J 1962 Knowledge and Belief: An Introduction into the Logic of the Two Notions, Ithaca: Cornell University Press.
- Hintikka J 1963 The Modes of Modality, Acta Philosophica Fennica 16, 65-82.
- Hintikka J 1967a Individuals, Possible Worlds, and Epistemic Logic, *Noûs* 1, 33-62.
- Hintikka J 1967b Existence and Identity in Epistemic Contexts, *Theoria* 33/2, 138–148.
- Hintikka J 1969, Models for Modalities, Dordrecht: Reidel.
- Hintikka J 1970 The Semantics of Modal Notions and The Indeterminacy of Ontology, *Synthese* 21/3, 408–424.
- Hintikka J 1971 Some Main Problems of Deontic Logic, Deontic Logic: Introductory and Systematic Readings, ed. by R. Hilpinen, Dordrecht: Reidel.
- Hintikka J 1973 Carnap's Heritage in Logical Semantics, Synthése 25, 372–397.
- Hintikka J 1973 Logic, Language-Games and Information: Kantian Themes in Philosophy of Logic, Oxford: Oxford University Press.
- Hintikka J 1974 Knowledge by Acquaintance Individuation by Acquaintance, Bertrand Russell: A Collection of Critical Essays, ed. by D. F. Pears, Anchor Books, Garden City (1972). (as reprinted in J. Hintikka 1974, Knowledge and the Known: Historical Perspectives in Epistemology, Dordrecht: Reidel (1974))
- Hintikka J 1975 The Intentions of Intentionality, Dordrecht: Reidel.
- Hintikka J 1980a On Sense, Reference, and the Objects of Knowledge, *Epistemologia* 3, 143 – 162, as reprinted in Hintikka & Hintikka (1989).
- Hintikka J 1980b Degrees and Dimensions of Intentionality, *Language, Logic, and Philosophy*, ed. by Haller & Grassl, Vienna: Hölder-Pichler-Tempsky. 69 82, as reprinted in Hintikka & Hintikka (1989).
- Hintikka J 1981 Kant's Theory of Mathematics Revisited, *Philosophical Topics* 12/2, 201-215.
- Hintikka J 1982 Is Alethic Modal Logic Possible, in Niiniluoto & E. Saarinen (1982), 1–15.
- Hintikka J 1983 Situations, Possible Worlds, and Attitudes, *Synthése* 54, 155–162.

- Hintikka J 1984 Kant's Transcendental Method and His Theory of Mathematics, *Topoi*, 3/2, 99-108.
- Hintikka J 1990 Cartesian Cogito, Epistemic Logic and Neuroscience: Some Surprising Connections, Synthese 83,133–157.
- Hintikka J 1996 World Lines and Their Role in Epistemic Logic, *Philosophical Logic and Logical Philosophy*, ed. by Bystrov & Sadovsky, Dordrecht: Kluwer.
- Hintikka J 1998 Perspectival Identification, Demonstratives and "Small Worlds", Synthése 114, 203–232.
- Hintikka J 1999 The Emperor's New Intuitions, *Journal of Philosophy* 96/3, 127–147.
- Hintikka J 2006 Intellectual Autobiography, in Auxier & Hahn (2006).
- Hintikka J 2007 Socratic Epistemology, Cambridge, Cambridge University Press.
- Hintikka J 2013 Epistemologia vetus et nova, Open Problems in Epistemology (Acta Philosophica Fennica 90), ed. by J. Hintikka, Helsinki: Philosophical Society of Finland.
- Hintikka J & Hintikka M B 1989 The Logic of Epistemology and the Epistemology of Logic: Selected Essays, Dordrecht: Kluwer.
- Hintikka J & Kulas J 1985 Anaphora and Definite Descriptions Two Applications of Game-Theoretical Semantics, Dordrecht: Reidel.
- Hintikka J & Sandu G 1995 The Fallacies of the New Theory of Reference, *Synthese* 104, 245–283.
- Hintikka J & Symons J 2003 Systems of Visual Identification in Neuroscience: Lessons from Epistemic Logic, *Philosophy of Science* 70, 89 104.
- Holliday W & Perry J 2014 Roles, Rigidity, and Quantification in Epistemic Logic, Outstanding Contributions in Logic 5: Johan van Benthem on Logic and Information Dynamics ed. by Baltag & Smets, Springer.
- Hughes G E & Cresswell M J 1996 A New Introduction to Modal Logic, New York, N.Y.: Routledge.
- Humphreys P W & Fetzer J H (eds.) 1998 The New Theory of Reference: Kripke, Marcus and its Origins, Dordrecht: Kluwer
- Jackendoff R 1972 Semantic Interprtation in Generative Grammar, Cambridge, Mass., MIT Press.
- Jackson F 1998a From Metaphysics to Ethics: A Defence of Conceptual Analysis, Oxford: Oxford University Press.
- Jackson F 1998b Reference and Description Revisited, *Philosophical Perspectives* 12, 201–218.
- Jackson F 2004 Why We Need A-Intensions, *Philosophical Studies* 118, 257 277.
- Jackson F 2011 Possibilities for Representation and Credence: Two Space-ism versus One Space-ism, in Egan & Weatherson (eds.) 2011.
- Jeffrey R C 1967 Formal Logic: Its Scope and Limits, New York: McGraw-Hill.
- Kaplan D 1968 Quantifying In, Synthése 19, 178-214.

- Kaplan D 1989 Demonstratives, *Themes from Kaplan*, ed. by Almog, Perry & Wettstein, New York: Oxford University Press.
- Kaplan D 1995 A Problem in Possible-Worlds Semantics, Modality, Morality and Belief, ed. by Sinnott-Armstrong et al, Cambridge: Cambridge University Press.
- Katz J 1966 The Philosophy of Language, New York: Harper.
- Katz J 1972 Semantic Theory, New York: Harper.
- Klima G (trans.) 2001 John Buridan *Summulae de Dialectica*, New Haven: Yale University Press.
- Knuuttila S 2008 Medieval Modal Theories and Modal Logic, Handbook of The History of Logic, Volume 2, Medieval and Renaissance Logic, ed. by Gabbay & Woods, Amsterdam: Elsevier.
- Kölbel M 2013 The Conversational Role of Centered Contents, *Inquiry* 56, 97–12.
- Kölbel M 2018 Perspectival Representation and Fallacies in Metaethics, Canadian Journal of Philosophy 48 (3-4), 379–404.
- Koskinen I, Limnell P & Vuorio T (eds.) 1996 *Luonto toisena, toinen luontona*, Tampere: University of Tampere.
- Kratzer A 2012 *Modals and Condotionals: New and Revised Perspectives*, Oxford: Oxford University Press.
- Kretzmann N & Stump E (trans.) 1988 William of Heytesbury, "The Compounded and Divided Senses" and "The Verbs 'Know' and 'Doubt", The Cambridge Translations of medieval Philosophical Texts, Vol. 1: Logic and Philosophy of Language, Cambridge: Cambridge University Press.
- Kripke S 1959a Semantical Analysis of Modal Logic (Abstract), Journal of Symbolic Logic 24/1, 323–324.
- Kripke S 1959b A Completeness Theorem in Modal Logic, *Journal of Symbolic Logic*, 24/1, 1–14.
- Kripke S 1963a Semantical Analysis of Modal Logic I, Zeitschrift f. math. Logik und Grund. d. Math., vol. 9, 67–96.
- Kripke S 1963b Semantical Considerations on Modal Logic, *Acta Philosophica Fennica* 16, 36–58.
- Kripke S 1977 Speaker's Reference and Semantic Reference, as reprinted in Kripke 2011.
- Kripke S 1980 Naming and Necessity, Cambridge, Mass.: Harvard University Press.
- Kripke S 2011 Philosophical Troubles, Collected Papers 1, New York: Oxford University Press.
- Kripke S 2015 Quantified Modality and Essentialism, Noûs 50/1.
- Lepore E & Smith B (eds.) 2006 Oxford Handbook of Philosophy of Language, Oxford: Oxford University Press.
- Lewis D 1972 General Semantics, Semantics of Natural Language ed. by D. Davidson & G. Harman, Dordrecht: Reidel.

- Lewis D 1979 Attitudes De Dicto and De Se, Philosophical Review, 88/4, 513-543.
- Lewis D 1983 Individuation by Acquaintance and by Stipulation, *Philosophical Review* 92, 3–32.
- Lihoreau F & Rebuschi M 2014 Epistemology, Context, and Formalism, Cham: Springer.
- Lindström S & Segerberg K 2007 Modal Logic and Philosophy, in Blackburn et al (2007).
- Logica Universalis 13/2, June 2019, (Special Issue: Hintikka Memorial Issue, Issue Editor: A. V. Pietarinen).
- Lycan W 1994 Modality and Meaning, Dordrecht: Kluwer.
- Martí G 2003 The Question of Rigidity in New Theories of Reference, NOÚS 37/1, 161–179.
- Moss S 2012 Updating as Communication, *Philosophy and Phenomenological Research* 85, 225–248.
- Niiniluoto I & Saarinen E (eds.) 1982 Intensional Logic: Theory and Applications (Acta Philosophica Fennica 35), Helsinki: Philosophical Society of Finland.
- Niiniluoto I 1979 Knowing that One Sees, Essays in Honor of Jaakko Hintikka ed. by E. Saarinen et al, Dordrecht: Reidel, 249–282.
- Niiniluoto I 1982 Remarks on the Logic of Perception, in Niiniluoto & Saarinen (1982), 183–202.
- Ninan D 2010 De Se Attitudes, Philosophy Compass 5, 551-67.
- Ninan D 2013 Self-Location and Other-Location, *Philosophy and Phenomenological Research* 87, 301–331
- Pagin P 2016 De Se Communication Centered or Uncentered?, in García-Carpintero & Torre (2016).
- Partee B 1975 Montague Grammar and Transformational Grammar, *Linguistic Inquiry* 6, 203–300.
- Peirce C S 1931–35 *Collected Papers of Charles Sanders Peirce*, vols. I–VI, ed. by Hartshorne and Weiss, Cambridge, MA, Harvard University Press.
- Perälä M 2014 Psychology of Language: Ancient and Early Medieval Theories, Sourcebook for the History of the Philosophy of Mind: Philosophical Psychology from Plato to Kant ed. by Knuuttila & Sihvola, Dordrecht, Heidelberg, New York, London: Springer-Verlag, 359–378.
- Perry J 1979 The Problem of the Essential Indexical, Noûs 13, 3–21.
- Perry J 2002 *Identity, Personal Identity, and The Self*, Indianapolis/Cambridge: Hackett Publishing Company.
- Perry J 2009 Hintikka on Demonstratives, Revue international de philosophie 4(250), 369 382.
- Pietarinen A V 2010 How to Analyse Rigid Designation? World Lines and Imperfect Information, *Proceedings of the Second ILCLI International Workshop on Logic and Philosophy of Knowledge, Communication and Action*, ed. by Arrazola, Xabier & Mara Ponte, San Sebastian: University of the Basque Country Press, 351–369.

- Plato Euthydemus, Plato: Laches, Protagoras, Meno, Euthydemus, (Loeb Classical Library, no. 165) (Greek and English Edition), trans. by W.R.M. Lamb, Cambridge, MA: Harvard University Press 1977.
- Prior A N 1955 Formal Logic, Oxford, Clarendon Press.
- Prior A N 1956 Modality and Quantification in S5, *Journal of Symbolic Logic* 21, 60–62.
- Prior A N 1971 Objects of Thought, Oxford: Oxford University Press.
- Putnam H 1975 Meaning of Meaning, Minnesota Studies in the Philosophy of Science 7, 131–193.
- Quine W V O 1956 Quantifiers and Propositional Attitudes, *Journal of Philosophy* 53, 177–187.
- Rantala V & Virtanen A 1996 Kuka keksikään Kripke-semantiikan, in Koskinen et al (1996).
- Recanati F 2007 Perspectival Thought, Oxford: Oxford University Press.
- Recanati F 2016 Indexical Thought: The Communication Problem, in García-Carpintero & Torre (2016).
- Russell B 1905 On Denoting, Mind, New series XIV, 479–493.
- Russell B 1912 The Problems of Philosophy, Oxford: Oxford University Press.
- Russell B 1918a Knowledge by Acquaintance and Knowledge by Description, *Mysticism and Logic and Other Essays*, London: Longmans Green & CO.
- Russell B 1918b The Philosophy of Logical Atomism, *The Monist* 28, 495-527, *The Monist* 29, 190-222, 345-380.
- Russell B 1919 Introduction to Mathematical Philosophy, London, Allen & Unwin.
- Sainsbury M 2014 Fishy Business, Analysis 74 (1), 3–5.
- Salmon N 1986 Frege's Puzzle 2nd ed, Atascadero: Ridgeview Publishing Company 1986/1991.
- Salmon N 2005 Are General Terms Rigid? Linguistics and Philosophy 28, 117-134.
- Sandgren A 2018 Secondary Belief Content, What is it Good For?, *Philosophical Studies* 175 (6), 1467–1476.
- Sandu G 2006 Hintikka and the Fallacies of the New Theory of Reference, *The Philosophy of Jaakko Hintikka*, ed. by Auxier & Hahn, Chicago, Open Court.
- Sandu G 2015 *Logic, Language and Games,* Acta Philosophica Fennica 91, Helsinki: Societas philosophica Fennica.
- Sandu G & Tanninen T 2017 Birth of Modal Logic in Northern Europe, Bulletin of Taras Shevchenko National University of Kyiv. Philosophy 2/2017, 9–19.
- Schiffer S 1987 Remnants of Meaning, Cambridge, Mass.: The MIT Press.
- Sintonen M 1994 Knowing and Making: Kantian Themes in Hintikka's Philosophy, *Investigating Hintikka* (*Grazer Philosophische Studien* 49), ed. by Haller, Amsterdam, Rodopi.
- Smith Q 1998 More Comprehensive History of The New Theory of Reference, in Humphreys & Fetzer (1998).

- Smullyan R 1968 First-Order Logic, New York: Springer-Verlag.
- Soames S 2002 Beyond Rigidity, Oxford: Oxford University Press.
- Soames S 2011 Kripke on Epistemic and Metaphysical Possibility: Two Routes to the Necessary *A Posteriori*, in Berger (2011).
- Sosa D 1996 The Import of the Puzzle about Belief. *Philos. Rev.* 105/3, 373–402.
- Stalnaker R 1972 [Review of Hintikka (1969)], Journal of Philosophy 69.
- Stalnaker R 1976 Propositions, *Issues in the Philosophy of Language*, ed. by MacKay & Merril, New Haven: Yale University Press.
- Stalnaker R 1981 Indexical Belief Synthése 49 (1), 129–151.
- Stalnaker R 2003a Conceptual Truth and Metaphysical Necessity", in Stalnaker 2003b.
- Stalnaker R 2003b Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays, Oxford: Oxford University Press.
- Stalnaker R 2004 Assertion Revisited: on The Interpretation of Twodimensional Modal Semantics, *Philosophical Studies* 118, 299–322.
- Stalnaker R 2016 Models and Reality, Canadian Journal of Philosophy 46/4–5, 709–726.
- Stanley J 2014 Constructing Meanings, Analysis 74/4, 662-676.
- Tanninen T 2019 Varieties of Rigidity, Logica Universalis, 13/2, 219-240.
- Thau M 2002 Consciousness and Cognition, Oxford, Oxford University Press.
- Torre S 2010 Centered Assertion, *Philosophical Studies* 150, 97–114.
- Torre S 2016 De Se Thought and Communication: An Introduction, in García-Carpintero & Torre 2016.
- Tselishchev V 1978 Hintikka's Possible Worlds and Rigid Designators, *Essays on Mathematical and Philosophical Logic*, ed. by Hintikka, Niiniluoto & Saarinen, Dordrecht: Reidel.
- Tulenheimo T 2009 Remarks on Individuals in Modal Contexts, Revue Internationale de Philosophie 63/250, 383–394.
- Tulenheimo T 2017 Objects and Modalities A Study in the Semantics of Modal Logic, Dordrecht: Springer.
- van Ditmarsch H & Sandu G (eds.) 2018 Jaakko Hintikka on Knowledge and Game-Theoretical Semantics, Outstanding Contributions to Logic 34, Cham: Springer.
- von Wright G H 1951a An Essay in Modal Logic, Amsterdam: North-Holland Publishing Company.
- von Wright G H 1951b Deontic Logic, Mind 60, 1–15.
- von Wright G H 1963 Norm and Action: A Logical Inquiry, Routlegde and Kegan Paul.
- Weatherson B & Egan A 2011 Introduction: Epistemic Modals and Epistemic Modality, in Egan & Weatherson (2011).
- Weber C 2013 Centered Communication, *Philosophical Studies* 166(1), 2015-2023.