

**Logical Consequence:  
Between Formal and Natural Language**

Thesis submitted for the degree of  
“Doctor of Philosophy”

by  
Gil Sagi

Submitted to the Senate of the Hebrew University

June, 2013

This work was carried out under the supervision of  
Professor Oron Shagrir and Professor Stewart Shapiro

## ABSTRACT

This dissertation is concerned with logical consequence and the idea that it is a formal relation. Logical terms are considered by many the key to the notion of form, and it is generally assumed that there is a strict division of terms into those that are logical (which are completely fixed in systems of logic) and those that are nonlogical (and are non-fixed). One of the main claims of this work is that this contention is false. Form, on the offered view, is determined by a system of “semantic constraints” which fix terms in different ways and to various degrees.

The dissertation has three parts. The point of departure is the discussion on logical terms in contemporary philosophy of logic. Part I is thus devoted to a critical survey of proposed criteria for logical terms. An extensive chapter provides a novel criticism of Gila Sher through a comparison with Stewart Shapiro. Sher is known for her isomorphism-invariance criterion for logical terms. Shapiro as well offers a condition on logical terms. It is proved that this condition is mathematically equivalent to Sher’s criterion. The mathematical equivalence provides a basis for comparison of Sher and Shapiro’s philosophical views. Another chapter is devoted to a survey of other writers proposing invariance criteria for logical terms (McGee, Feferman and Bonnay). The final chapter of Part I is concerned with criteria for logical terms which employ epistemic notions, advocated by Peacocke and McCarthy.

Part II presents the novel framework of semantic constraints using model-theoretic foundations: semantic constraints are defined as restrictions on admissible models. First, the basic ideas and philosophical motivation for the framework are brought forth. Next, an extensive chapter is devoted to the formal details of the framework. The new setting allows for novel definitions of various concepts such as dependency, determinacy and categories of terms, which are presented with relevant results. Part II also contains a detailed discussion of the notion of fixing a term in a formal system. In addition, there is a chapter that extends

the ideas of previous chapters to systems of modal logic.

Part III deals with the question of logical consequence in natural language. The main claim in this part is that natural language does not have a relation of logical consequence, at least as this relation is traditionally conceived in the philosophy of logic. This claim is backed by a historical and conceptual analysis. However, it is argued that logic can be imposed on natural language through the process of formalization, which involves commitments made with respect to the use and interpretation of a language. Here the discussion relates to previous sections, since commitments may come in the form of semantic constraints as those were presented earlier.

*Dedicated with love to the memory of my grandmothers,  
Irena (Henu) Weissberger (1922 Košice - 2012 Kfar Sava)  
and Zippora Argov (1923 Vienna - 2011 Nahariya).*



# Acknowledgments

I have been fortunate to be able to devote four years of doctoral studies to research in the philosophy of logic. This was made possible by the Polonsky fellowship I received from the Hebrew University, for which I am grateful.

During those years I have benefited from the help and support of numerous people. First and foremost, I am indebted to my supervisors Oron Shagrir and Stewart Shapiro. Oron Shagrir was there for me every step of the way with invaluable guidance and advice, and I cannot imagine carrying out this project without his support. Stewart Shapiro has agreed, to my great fortune, to work with me as an external supervisor. I am deeply thankful for all the inspiration he has provided to me in writing, in teaching and in conversation.

I wish to express my gratitude to the members of my committee Carl Posy and Eli Dresner. I have greatly benefitted from conversations with Carl Posy, who was also my teacher at various stages of my studies. I am thankful to my friends and colleagues at the Hebrew University, Ron Aboodi, Mordechai Applebaum, Aviv Keren, Eli Pitcovski and Jonathan Yaari, for long sessions of feedback and discussion. I am especially indebted to Ran Lanzet, whose comments have had a major effect on the contents of this work.

During my studies, I have spent considerable periods of time at the University of St Andrews as a visitor to the research project *Foundations of Logical Consequence*. In St Andrews I was first exposed to the wonderful philosophy of logic international community. Stephen Read, the leader of the project, has been a fantastic host and teacher to me. And

so have been other members of the project: Colin Caret, Michael De, Noah Friedman-Biglin, Ole Hjortland, Toby Meadows and Julien Murzi, who besides providing generous professional support, have also become my very close friends. I am also thankful to Hannes Leitgeb, Greg Restall and Gila Sher with whom I had the opportunity to discuss my work and so benefit from their penetrating comments.

Finally, I would like to thank my family: my parents Michal and Dov Sagi, and my brother Yaniv Sagi for their endless love and constant support.



## Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>I</b>	<b>Logical Terms</b>	<b>7</b>
<b>2</b>	<b>Introduction</b>	<b>9</b>
2.1	From Logical Consequence to Logical Terms	10
<b>3</b>	<b>Models and Logical Consequence</b>	<b>17</b>
3.1	Introduction	17
3.2	Necessity and Formality	20
3.3	Models and Logical Consequence	23
3.4	What Models Represent	27
3.4.1	Metaphysical Semantics	27
3.4.2	Linguistic Semantics	29
3.4.3	The Formal-Structural View	32
3.4.4	The Blended Approach	39
3.5	Tying the threads	43
3.6	Conclusion	45
3.7	Appendix	46

## TABLE OF CONTENTS

<b>4</b>	<b>Varieties of Invariance</b>	<b>49</b>
4.1	Invariance Criteria for Logical Terms: The Basics	49
4.1.1	Permutation invariance	50
4.1.2	Isomorphism invariance	51
4.2	McGee's Justification of Permutation Invariance	54
4.2.1	Summary	54
4.2.2	Connectives and Operations	55
4.2.3	Tarski's thesis	56
4.2.4	The argument from $L_{\infty\infty}$	57
4.2.5	Concluding remarks	65
4.3	Feferman: Invariance under Homomorphisms	67
4.3.1	Summary	67
4.3.2	Against the Tarski-Sher thesis	67
4.3.3	Feferman's criterion	69
4.3.4	The modified criterion and definability in FOL	71
4.3.5	The problematic use of formalisms in the philosophical debate on logicality	73
4.3.6	Concluding remarks	77
4.4	Bonnay: Potential Isomorphisms	79
4.4.1	Summary	79
4.4.2	Bonnay's criterion	79
4.4.3	Generality	83
4.4.4	Formality	85
4.4.5	Concluding remarks	86
<b>5</b>	<b>Epistemic Criteria for Logical Terms</b>	<b>87</b>
5.1	The Epistemic Component in Logical Consequence	87

## TABLE OF CONTENTS

5.2	Peacocke's Criterion for Logical Terms	88
5.3	Criticisms of Peacocke's Criterion	92
5.4	The Modal and Epistemic Arguments against the Invariance Criterion and McCarthy's Modification	96
5.5	Concluding Remarks	102
<b>II</b>	<b>From Logical Terms to Semantic Constraints</b>	<b>105</b>
<b>6</b>	<b>Formality in Logic: From Logical Terms to Semantic Constraints</b>	<b>107</b>
6.1	Introduction	107
6.2	Logical Terms	108
6.3	Semantic Constraints	111
6.4	Meaning postulates, Cross-Term Restrictions and Semantic Constraints	117
6.5	Conclusion	120
<b>7</b>	<b>Semantic Constraints: The Framework</b>	<b>123</b>
7.1	Language, Models, and Semantic Constraints	124
7.1.1	The language L	124
7.1.2	Models	124
7.1.3	Semantic constraints	124
7.1.4	The status of the domain	125
7.2	Examples of Constraints	125
7.2.1	Propositional logic	125
7.2.2	Standard first order predicate logic	126
7.2.3	Constrained FOL	128
7.3	Basic Notions in the Framework	130
7.3.1	Constraining a phrase and non-trivial constraints	130

## TABLE OF CONTENTS

7.3.2	Sentences and singular phrases	131
7.3.3	Relations between terms: determinateness and dependency	132
7.3.4	Categories and schemas	139
7.3.5	Families of terms	144
7.3.6	Families: Examples	148
7.3.7	Equivalence	151
7.4	A Remark on Proof Theory	152
7.5	Appendix: p-categories	155
<b>8</b>	<b>Determinacy, Invariance and Fixing a Term</b>	<b>159</b>
8.1	Fixing a Term	160
8.2	Fixity and Determinacy	175
8.3	Invariance and Determinacy	179
8.4	Conclusion	185
<b>9</b>	<b>Semantic Constraints: Modal Logic</b>	<b>187</b>
9.1	Modal Systems	188
9.1.1	Standard modal systems	188
9.1.2	Other modal operators	190
9.1.3	Determinacy, dependence and categories	192
9.2	Determinacy, Invariance and Fixing a Term: the Modal Logic Case	194
9.2.1	Fixing terms and modal realities	196
9.2.2	Is <i>Box</i> a logical term?	204
9.2.3	Modal logic and degrees of logicity	208
9.3	Conclusion	210

*TABLE OF CONTENTS*

<b>III</b>	<b>Logic and Language</b>	<b>213</b>
<b>10</b>	<b>Logical Consequence and Natural Language</b>	<b>215</b>
10.1	Is there logical consequence in natural language?	217
10.1.1	Making sense of the question: historically	218
10.1.2	Making sense of the question: conceptually	229
10.1.3	Logic as methodology and natural language	231
10.1.4	Conclusion	241
10.2	Logic as a system of commitments	242
10.3	Revisiting logic-as-model	249
<b>11</b>	<b>Conclusion</b>	<b>255</b>

# Chapter 1

## Introduction

Logic holds an enticing promise: Some claims can be made on the basis of others in a completely secure manner purely based on the forms of the claims involved. The logical tradition has customarily concerned itself with *formal validity*, and it offers the paradigmatic example:

All Greeks are human

All humans are mortal

Therefore, All Greeks are mortal

In basic logic classes it is said that one can see from the “form” of the argument alone that if the premises are true, so must be the conclusion. It is further explained that “Greeks”, “human” and “mortal” can be substituted by any other words of the same syntactic category and the argument will remain valid: the argument is therefore formally valid. The other words in the argument constitute its form. When an argument is checked for its formal validity, the form is held fixed, and all the rest is allowed to vary.

Not all accounts of validity are committed to the formality of logic, and not all those committed to the formality of logic speak of forms of arguments. Nonetheless, the idea that logical validity is a matter of the form of arguments is dominant throughout the history of

the subject. We find this idea of validity already in Aristotle, even if not articulated in the same terms: Aristotle did not speak of forms of arguments and formal validity (MacFarlane, 2000). However, the method of substitution is central to his theory of the syllogism. And holding some of an argument fixed while letting the rest vary is common lore to this day.

Contemporary philosophers of logic approach the issue of the forms of sentences and arguments through the notion of a *logical term*. The terms of a language are divided into two types of terms: logical and nonlogical. It is generally assumed that the form of a sentence is determined by its grammatical structure and the logical terms it contains. Thus the question of *which* terms are logical has become central in the literature on logical validity. In the argument above, “all” and “are” would be considered logical (rendered into logical formalism, the logical terms of the argument become  $\forall$  and  $\rightarrow$ ). A general theory of formal validity thus needs to answer what makes these and other terms logical.

The assumption that logical terms have a distinctive role in determining forms of sentences is rarely contested. In this work, the generally assumed centrality of logical terms is put into question. I propose a new framework for logical systems where the terms of a language are not strictly divided into logical and nonlogical. I still accept that logic is formal, but not the conventional understanding of form. Form, in the proposed alternative framework, is perceived as set by *semantic constraints*: restrictions on a language which fix terms in various manners and to different degrees. A semantic constraint can also fix the mutual interpretation of some terms. An example of such a constraint would be: ‘the extension of *all – green* and *all – red* are mutually exclusive’. In the new framework, all terms that are not completely fixed are still allowed to vary, but their freedom of variation is not set by their syntactic category, but rather by the constraints imposed on them.

The idea of semantic constraints is developed in this work through a formal framework in which various systems of constraints can be construed. A more philosophical outlook is proposed in a discussion on logic and natural language, where I propose to view semantic constraints as commitments to a linguistic framework made by participants of a discourse.

## CHAPTER 1. INTRODUCTION

\* \* \*

The concept of logical consequence, at the heart of this dissertation, is one of the age-old subjects of philosophical inquiry. The aim of this work is not to provide a theory of the historical or current use of this concept. Nor does this work aim to capture a Platonic Idea that might guide the use of this concept. The aim of this investigation is to look into the basic motivations guiding the use of the concept, and contribute to its explication on the basis of these motivations. I find that the requirement that logic be formal has strong motivations. One of the basic motivations for this requirement, I suggest, is that formality helps avoid contentious metaphysical questions and places reasoning on secure grounds. This motivation does not enforce a strict division of a language into logical and nonlogical terms, and from here emanates the theory of semantic constraints.

This work does not start from scratch: it is firmly situated in the Tarskian tradition in logic. Most works I discuss are in this tradition. It might be useful to point out the elements of the Tarskian tradition which have a major role in this work.

The new proposed framework uses the Tarskian machinery of model-theoretic semantics. The distinction between logical and nonlogical terms, which is deeply rooted in the Tarskian tradition, is not committed to. However, my alternative understanding of form is still, I believe, very much in the spirit of this tradition.

Two main principles of the Tarskian tradition are endorsed in this work:

*Necessity:* In a logically valid argument, the conclusion follows necessarily from the premises, or, equivalently, it is not possible for all premises to be true and the conclusion false.

*Formality:* The logical validity of an argument is determined by its form. Specifically, an argument is valid if and only if every other argument with the same form is valid.



## CHAPTER 1. INTRODUCTION

These principles, inspired by Tarski (1936) are presented and discussed in Ch. 3, but are pertinent to almost all other chapters of this work. *Formality* is the main focus of this work. *Necessity* will be mostly assumed without critical evaluation. However, some ideas about *Necessity* are germane to this work. When understood through the notion of possible worlds, *Necessity* becomes: In every possible world in which the premises are true, so is the conclusion. The Tarskian definition of logical consequence states that an argument is valid if in every model in which the premises are true, so is the conclusion. It is thus very natural to take models, which are set-theoretic objects used in a mathematical definition, to represent possible worlds. This idea, which is brought up and criticized by Etchemendy, is discussed in Ch. 3. Modifications of this idea are endorsed throughout this work, and lay the basis of understanding the relation between the mathematical definition of logical consequence and the philosophical concept it aims to capture.

\* \* \*

This work consists of three parts. Part I is devoted to a critical survey of various approaches to logical terms. A major chapter in this part (Ch. 3, based on (Sagi, forthcomingb)) compares the approaches of Gila Sher and Stewart Shapiro to logical consequence. The main point of comparison is the role Sher and Shapiro attribute to models in the Tarskian definition of logical consequence. However, the issue of logical terms is crucial to both of the views considered. I show that the mathematical conditions Sher and Shapiro give for logical terms are equivalent, and evaluate their philosophical positions based on this common background. In other chapters of Part I, I discuss a variety of approaches to logical terms. I focus on two types of criteria for logical terms that have been proposed in the literature. One is that of “invariance criteria” for logical terms, by which logical terms are invariant under certain kinds of transformations of the domain. Specifically, I discuss McGee, Feferman and Bonnay (Ch. 4). Sher, discussed mainly in Ch. 3, is a primary proponent of invariance criteria. In addition, I discuss “epistemic criteria” for logical

## CHAPTER 1. INTRODUCTION

terms: criteria that involve the a priori and other epistemic notions. I concentrate on the work of Peacocke and McCarthy. I devote the main part of the discussion to a criticism of McCarthy’s objection to invariance criteria (Ch. 5).

Part II is devoted to the new framework of semantic constraints. I first provide the philosophical motivation for moving from a term-based framework (where terms are strictly divided into logical and nonlogical) to a framework of semantic constraints (Ch. 6). Next, I build up the framework formally. I define and prove some results with respect to relevant concepts such as *dependence*, *determinacy* and *categories* of terms. Since the new framework is more general than the conventional term-based framework, logical terms can be defined as well as special cases of completely fixed terms (Ch. 7). I then return to some open questions about logical terms. It is common to characterize logical terms as those terms that are “held fixed” in a system, but the notion of being fixed is usually assumed. Using the machinery developed for the new framework, I further develop the definition given to fixing a term in Ch. 7, and provide a novel explication of what is to “fix” a term. The insights about fixing terms yield also a way of evaluating which terms *should* be fixed, or of measuring the logicity of terms. Generalizing to the framework of semantic constraints, the same insights can be employed to evaluate which constraints rather than terms should be fixed in a system (Ch. 8). Finally, I extend the framework to encompass systems of modal logic. Moreover, I argue that the framework of semantic constraints is better suited than the term-based one for systems of modal logic. The notion of fixing a term which was dealt with in Ch. 8, is further developed, as it is subjected to new complications in the modal framework (Ch. 9).

Part III consists of one chapter (Ch. 10.1), and concerns the relation between logical consequence and natural language. In the main section of this chapter I argue that unregulated natural language does not have a relation of logical consequence (Section 10.1). I devote a large part of this section to an explanation of what it is for natural language to have a relation of logical consequence, motivated both historically and conceptually. Next,

## CHAPTER 1. INTRODUCTION

I propose a more positive outlook, and claim that a natural language can have a relation of logical consequence once it is formalized. I take up a Carnapian position towards linguistic frameworks, in combination with the new idea of semantic constraints. I employ the notion of form developed in Part II, which was explicated through the notion of semantic constraints. In the process of formalization, semantic constraints are imposed on a language as explicit commitments of reasoners or participants of a discourse (Section 10.2). Finally, I discuss some implications of my position through its comparison with Shapiro's *logic-as-language* approach (Section 10.3).

## Part I

# Logical Terms



## Chapter 2

# Introduction

In this part of the dissertation I survey the main contemporary approaches to logical terms. I concentrate on two classes of criteria for logicity. The first class is that of “invariance criteria”, that rely on a model-theoretic notion of invariance under structures. The main contributors to this approach are Tarski, McCarthy, Sher, McGee, Feferman and Bonnay. Stewart Shapiro proposes a necessary condition for logical terms that is equivalent to Sher’s invariance criterion. I devote a chapter to a detailed comparison of Sher’s approach with Shapiro’s.

The second class of criteria I discuss is that of “epistemic criteria”, that incorporate the notion of *a priori*. Here the main figures are Peacocke and McCarthy. This chapter is slimmer. It touches upon issues that cannot be dealt thoroughly in the limits of this dissertation (namely, the notion of *a priori*), and on issues that will be dealt with more extensively in the part on logic in natural language (namely, the issues of analyticity and of form in natural language). My goal is to provide a faithful yet critical presentation of the epistemic approaches, without going too far afield.

The two classes of criteria I deal with are not mutually exclusive. Some criteria explicitly incorporate both epistemic and invariance notions, such as McCarthy’s. In addition, some

proponents of invariance criteria, such as Sher, rely on epistemic notions at the motivational level.

In an attempt to provide a comprehensive picture, I first describe shortly in the next section some alternative approaches to logical terms. A thorough presentation of various approaches is given by MacFarlane’s encyclopedic entry on “Logical Constants.” ([MacFarlane, 2009](#)).

## 2.1 From Logical Consequence to Logical Terms

The problem of logical terms arises upon the realization that the boundary between logical and nonlogical terms has a distinctive role in determining logical consequence. It is a commonly accepted assumption that logical consequence is a formal relation, and as such depends only on the forms of sentences. A form of a sentence, as commonly conceived, is determined by its grammatical structure and the logical terms involved. The logical terms are those whose meanings persist when only form is considered.

Bolzano is a precursor to the modern day semantic conception of formality. Bolzano spoke of propositions and ideas composing them. An idea can be made variable in a proposition, and then other ideas can be substituted to form other propositions. An *analytic proposition* is a proposition in which at least one idea can be made variable without altering the truth or falsity of the proposition. The *logically analytic propositions* constitute a subclass of the analytic propositions, and are such that no other than logical knowledge is required. According to Bolzano, in logical propositions, the ideas composing the invariant part all belong to logic. However, Bolzano is skeptical as to whether an uncontroversial distinction could be drawn between logical and nonlogical ideas ([Bolzano, 1973](#), §148).

Tarski advanced the semantic conception with his model-theoretic definition of logical consequence ([Tarski, 1936](#)), which will be discussed extensively in Chapter 3. However, when presenting the definition he left the question of logical terms open. The question is

## CHAPTER 2. INTRODUCTION

pertinent to logical inquiry since different choices of logical terms lead to different extensions of the relation of logical consequence.

Now take a first order language without identity. To make things simpler, take a first order language with negation, implication and the existential quantifier as the only logical terms. Which are the logical terms of this language? Indeed, they were just listed. The logical terms of formal languages are predefined as such. But notice that the phrase “logical term” is used in two ways in the logic literature. Sometimes it’s a system relative notion, meaning “the terms for which the interpretation is held fixed”.<sup>1</sup> This sense pertains to the functional role of logical terms in formal systems. And sometimes the ascription of logicity to a term is meant in a non-relative way, implying that the term satisfies some conditions that entitle it to be characterized as logical. Let us distinguish between *logical terms in the shallow sense* and *logical terms in the deep sense*. While logical terms in the shallow sense are stipulated in and relative to a system (they are simply defined by it as such), the notion of terms being logical in the deep sense, “by their nature”, needs further analysis and specification.<sup>23</sup>

Simply fixing a term will not make it logical by its nature (or so goes a common conception): the question of logical terms is which terms *should* get fixed and be considered

---

<sup>1</sup>That is, in a model-theoretic setting, logical terms are defined at the outset, and their interpretation in a model is a function of the domain of the model. See Chapter 8 for elaboration of this issue.

<sup>2</sup>This distinction between the deep and the shallow senses of *logical term* is in the spirit of Evans’s distinction between immanent and transcendent definitions in semantics. According to Evans,

One provides an immanent definition of some semantical term  $W$  if one does not define it absolutely but rather defines the notion  $\ulcorner e \text{ is } W \urcorner$  according to theory  $T$ . One provides a transcendent definition when the definition contains no such relativity to a theory; when one says, rather, what a theory *ought* to treat as  $W$ .

(Evans, 1976, p. 200).

<sup>3</sup>A similar distinction can be made from a proof-theoretic perspective, where logical terms understood in the shallow sense would be relative to a proof system, and understood in the deep sense they satisfy some criterion for logicity.



as logical – what is the correct criterion for choosing the terms that are to be fixed in the system. It will be useful to spell out the underlying assumption driving the search for a criterion for logical terms:

PD. There is a principled distinction between logical and nonlogical terms.

Logical systems have been used and employed without the question of logical terms being settled, and even before the question has gained prominence for theoreticians. In the standard presentations of logical systems, the logical terms are simply stipulated, without reference to a criterion. When logical terms are merely stipulated, only the shallow sense is employed. Logical terms then simply “come with” the system. Even after the problem of logical terms was explicitly raised, it has been suggested that logical terms should just be given by enumeration (White & Tarski, 1987; Quine, 1953). Of course, the stipulative method suffers from arbitrariness. The standard systems that were presented before through stipulation, first-order logic in particular, have proven to be very fruitful, and have instigated a search for a rationale that would explain the accepted group of logical terms, and would perhaps lead to a general criterion.

It is not always clear with all proposed criteria whether they are meant to characterize logical terms in the shallow sense or in the deep sense. Some criteria are system relative, but this does not entail that they are “shallow”. Carnap has characterized the logical terms such that each sentence constructed solely from them is determinate (that is, either logically true or logically false in the system, (Carnap, 1937, p. 177)). Carnap’s characterization appears to be shallow—it is a possible explication of the notion of a *fixed term*. However, it seems that Carnap is trying to track a deeper distinction between *logical* and descriptive terms, the former conceived of as mathematical and the latter as pertaining to empirical objects and properties. We are left with two salient options of understanding Carnap: either the way the system is built needs to conform to a prior distinction between the logical and the descriptive, or that distinction itself is system relative and thus conventional. The first

## CHAPTER 2. INTRODUCTION

option relates to logical terms a deep sense, and the second blurs the distinction between the deep and the shallow senses. Carnap, in (Carnap, 1948), is more hesitant about the definition of logical terms. He relates to the underlying intuition distinguishing logical from descriptive signs, but makes do with giving the primitive logical signs by enumeration for each system (p. 58). The relation to the *L-concepts* is then proved, and is not definitional of logical terms as in (Carnap, 1937). Carnap expresses skepticism as to whether a general definition can be given (p. 59). But then, in (Carnap, 1963a, pp. 931f) Carnap says that the distinction is easy to make, and that there is no real doubt regarding any given term. Other apparently shallow characterizations are given by Quine and Davidson (Davidson, 1984; Quine, 1986), the latter criticized by Evans (Evans, 1976). See (MacFarlane, 2009).

In recent years, various criteria based on a principled distinction between logical and nonlogical terms have been proposed. I discuss some of the main approaches in subsequent chapters. This work draws its outlook and methods from the Tarskian model-theoretic tradition, but it is important to note that the problem of logical terms arises in the proof-theoretic tradition as well. In the approach advanced by Gentzen (Gentzen, 1969, p. 80), logical terms are defined or characterized by rules. The demarcation between logical and nonlogical terms is done through local conditions on the terms' introduction and elimination rules, as well as more global conditions on the whole system of proof (such as conservativeness, the subformula property etc.) (see (Popper, 1947; Hacking, 1979; Hodes, 2004; Peacocke, 2004)) I shall not discuss proof-theoretic approaches, except for Peacocke's, in connection to his earlier model-theoretic oriented approach.

Alongside proposals for criteria for logical terms in recent literature (Peacocke, 1976; McCarthy, 1981; Sher, 1991; Warmbröd, 1999; Feferman, 1999; Bonnay, 2008), doubts have been raised as to the prospects of such a project.<sup>4</sup> It seems as if all proposed criteria are problematic and highly controversial. This does not yet prove that a correct criterion does not exist, but it would reasonably lead to skepticism in the project (Etchemendy, 1990;

---

<sup>4</sup>See (MacFarlane, 2009) for a survey and discussion of the various approaches to logical terms.

Dutilh Novaes, 2012) and give rise to relativistic approaches (Varzi, 2002; Tarski, 1936).

The notion of *logical term* is a curious case in the philosophy of logic. Since logical terms affect the extension of logical consequence and logical truth, it seems that they should be determined according to the properties of the latter concepts. However, it seems that the concept of logical terms has developed a life of its own. Thus, authors relate directly to properties of logical terms and intuitions regarding what could and what couldn't be a logical term, without reference to the concepts of logical consequence and logical truth on which they depend. Specifically, it is not clear why holding logical terms *fixed* should lead to the right extensions of logical consequence and logical truth. This issue is raised with respect to Peacocke in Chapter 5 (and see (Warmbrød, 1999) for a similar criticism). It is interesting to note that Tarski's definition of logical notions (Tarski, 1986) employs direct intuitions regarding logical notions as the subject matter of logical theory, but where the definition is given the concepts of logical consequence and logical truth are purposefully left out, and no claim regarding the connection of logical notions to those concepts is made. Even the connection between logical *notions* (i.e. the set-theoretic entities which Tarski discusses) and logical *constants* is not discussed (but see also (Tarski & Givant, 1987), where a connection between logical notions (there: *objects*) and logical constants is made). Indeed, it has been suggested that although there are various nice logical properties terms may have and for which they can be considered logical, these different properties need not coincide, and logical terms might not form a “natural kind” (van Benthem, 1989, p. 317). Another way to look at it is that *logical term* might not be univocal, and that different purposes call for different sharpenings of the concept. In that case, a proposed criterion should always be accompanied with the purpose in respect to which it is considered.

The view that logical consequence is relative to the set of logical terms has been termed *Tarskian Relativism* (Varzi, 2002), following the skepticism Tarski raised at the end of (Tarski, 1936):

## CHAPTER 2. INTRODUCTION

TR. Logical validity is relative to a choice of logical terms, and there is no principled distinction between logical and nonlogical terms.

Tarskian relativism employs the shallow meaning of “logical terms”. Yet, logical terms still play a major role: they are an important parameter that must be fixed for logical consequence to be determined.<sup>56</sup>

Another way of solving the problem of logical terms without accepting (PD) is through pragmatic considerations. Warmbrød distinguishes between *core* logic and *extended* logic (Warmbrød, 1999). Core logical theory provides a common ground for scientific theories, and therefore should be non-controversial and be based on minimal assumptions. The core theory thus incorporates only a minimal set of logical terms needed for formulating scientific theories. Extended logical theory is aimed for other uses of logic, such as modeling phenomena in natural language. Whereas core logic is limited to first-order logic without identity, extended logic may include the whole variety of modal and tense logics. In Warmbrød’s pragmatic approach is a brand of relativism: logicality is relative to a purpose.

Gómez-Torrente offers a pragmatic approach as well (Gómez-Torrente, 2002). He agrees that there may be necessary conditions on logicality, but casts doubt that a good mathematical necessary and sufficient condition that captures the relevant semantic and epistemic intuitions will come up. The considerations at play are of a complex pragmatic sort, having to do with specific utilities of logical systems. (For other approaches employing pragmatic considerations, see (Wagner, 1987; Hanson, 1997). See also (MacFarlane, 2009).)

Finally, another way of dismissing (PD) and the need for a definite criterion for logical

---

<sup>5</sup>We should note that in (Tarski, 1936) it is not suggested that any choice of logical terms would be as good as another, but merely that there is not one correct choice.

<sup>6</sup>Relativity of logical validity to a set of terms is sometimes attributed to Bolzano, he starts out with a relative notion of validity (more precisely: Bolzano spoke of the degrees of validity (Bolzano, 1973, §147)), but as was noted earlier, he also provides non-relative notions of analytic proposition and logically valid proposition.

terms is by rejecting the condition of *Formality*, that logical consequence depends on forms of sentences—at least in the usual way it is understood.<sup>7</sup> Such a view has been advanced by (Etchemendy, 1983; Read, 1994; Dutilh Novaes, 2012) (and see also (MacFarlane, 2009)). In this work *Formality* is assumed as a basic condition on logical consequence, alongside a condition of *Necessity* (see Chapter 3). There is a long tradition of defining logic as formal (see (MacFarlane, 2000)), and the condition of *Formality* stems from this tradition. The main motivation for *Formality* given in this work is that formality is a means to avoid metaphysical questions (see Chapter 3). But there is an understanding of logical consequence that accepts only *Necessity*, without *Formality*. Such an approach, it seems, aims at a somewhat different concept of *logical consequence* and *validity* than mine and the others mentioned above do.

---

<sup>7</sup>Those who deny *Formality* usually take it to entail a strict division of terms into those that are logical and those that are nonlogical. I present an alternative view of formality in Chapter 6 which shows that the doubts with respect to logical terms need not lead to a rejection of *Formality*.

## Chapter 3

# Models and Logical Consequence

### 3.1 Introduction<sup>1</sup>

Gila Sher and Stewart Shapiro each propose a formal criterion for logical terms within a model-theoretic framework, based on the idea of invariance under isomorphism. In this chapter, I prove that the two criteria are formally equivalent, and thus we have a common ground for evaluating and comparing Sher and Shapiro philosophical justification of their criteria. The main point of contention is what the models of model theory should be taken to represent. It is argued that Shapiro *blended approach*, by which models represent possible worlds under interpretations of the language, is preferable to Sher's *formal-structural view*, according to which models represent formal structures. The advantages and disadvantages of both views' reliance on isomorphism are discussed.

There is no dispute that Tarski's definition of logical consequence is one of the landmarks of logic in the 20th century. The model-theoretic definition Tarski devised has such a strong intuitive appeal, that for many philosophers and mathematicians today model-theoretic consequence just is logical consequence. Indeed, Tarski's accomplishment is not one of

---

<sup>1</sup>This chapter is an edited version of (Sagi, forthcomingb).

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

those earth-shattering surprising mathematical results; it has an air of obviousness, at least in retrospect – which is a main source of its strength.

Tarski intended to capture, with his model-theoretic definition, a pre-existing concept of consequence, in a manner as faithful as possible to its ordinary usage. Tarski analysed the ordinary concept of consequence as *necessary* and *formal* – and those are the features that he was committed to capturing. So, despite the general acceptance of Tarski’s definition, there is still room to ask whether, or how, it captures the main features of logical consequence.

The philosophical question of the basis of the model-theoretic definition of consequence, in its bluntest and most critical form, was formulated by John Etchemendy ([Etchemendy, 1990, 2006](#)). Etchemendy argued that Tarski’s definition is both conceptually and extensionally defective. Though Etchemendy’s attack was not received as detrimental to the model-theoretic project, it posed a two pronged challenge to its adherents. One part of the challenge was to defend Tarski by refuting Etchemendy’s interpretation, and another was to argue for the adequacy of model theory for defining consequence, regardless of the specifics of Tarski’s work.

The main focus of this chapter will be the second part of Etchemendy’s challenge. I will address the issue of supporting the model-theoretic definition through the works of two main proponents of the semantic tradition: Gila Sher and Stewart Shapiro. Different emphases in their work may have caused the impression that they are not contributors to the same debate, but I will show that the contrary is true: they are in fact dealing with the same issues and have much in common.

Two threads intertwine throughout this chapter. One is mathematical, and the other philosophical. The theorists we deal with share a common philosophical framework, which consists of an analysis of the concept of logical consequence as necessary and formal, and a method for showing that the mathematical apparatus adequately captures this concept. Their common method for showing the adequacy of model theory is to take models to represent something in the non-mathematical realm. The point of comparison here will

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

be the sort of items that models represent, and this is where Sher and Shapiro diverge. Sher takes models to represent *formally possible structures*, whereas Shapiro takes them to represent *possible worlds under interpretations of the nonlogical terms of the language*. I argue that Shapiro’s philosophical account is more successful: it captures necessity and formality in a straightforward manner which is faithful to the conceptual common ground.

The common mathematical framework will be a first order language with model-theoretic semantics. Within this framework I compare different ways of fixing the logical terms of the language. Both Sher and Shapiro provide mathematical criteria for logical terms which are essential to their accounts of formality. In the course of the chapter, I point out (and prove in the appendix) that their formal criteria for logical terms are equivalent, and this draws their approaches even closer. Their common mathematical framework allows a close comparison of their philosophical views.

The plan of the chapter is as follows. In section 2, I formulate and discuss the conditions of *Necessity* and *Formality* on logical consequence, in a way, I think, that would be agreeable to the philosophers I discuss. I refer to the notion of the *form* of a sentence, but do not provide a full analysis of it. I attribute to the form of a sentence a certain role: referring to it allows us to argue validly, while avoiding intricate metaphysical questions. This minimalistic characterization of form is compatible with the views I discuss, and will suffice for the evaluation of various accounts of model-theoretic consequence. In section 3, model theory enters the picture. I present Shapiro’s *logic-as-model* approach that sets the ground for the connection between the mathematical apparatus of model theory and the philosophical notion of logical consequence. In section 4, I map the different possible approaches within the *logic-as-model* framework, among which are Sher and Shapiro’s views. I go through some disadvantages in Sher’s view that Shapiro overcomes. In section 5, I point out, based on (Sagi, 2011), that Shapiro and Sher’s mathematical characterizations of logical terms are equivalent, and through an additional mathematical result, I substantiate the claim in favour of Shapiro’s philosophical account. Finally, following some concluding remarks in



section 6, the appendix in section 7 contains an outline of the proof of the equivalence result.

## 3.2 Necessity and Formality

I shall focus here on a fairly conventional understanding of logical consequence as *necessary* and *formal*. I will make only minimal assumptions on what necessity and formality amount to, and will not offer a full explication of these notions. In my assumptions, I follow by and large Tarski’s outline of the intuitive concept of logical consequence (Tarski, 1936, p. 414). This is by no means intended to be an interpretation of Tarski’s idea, but just a common ground to start from, general enough to be consistent with many of the formulations. Let an *argument* be a pair consisting of a set of sentences (the premises) and a sentence (the conclusion).<sup>2</sup> Inspired by Tarski, we can characterize logical consequence by two conditions:

*Necessity:* In a logically valid argument, the conclusion follows necessarily from the premises, or, equivalently, it is not possible for all premises to be true and the conclusion false.

*Formality:* The logical validity of an argument is determined by its form. Specifically, an argument is valid if and only if every other argument with the same form is valid.

Indeed, the two authors I focus on take on board these conditions. For Sher, “The intuitive notion of logical consequence is that of necessary and formal consequence” (Sher, 1996, p. 668). Sher has an elaborate conception of formality that goes beyond the condition of *Formality* I formulated above, but that Tarskian condition is certainly acceptable by her. Formality, she would add, is primarily attributed to properties of objects in the world, and only derivatively to items in a language, as the latter denote the former.

---

<sup>2</sup>For the purpose of this work I take the *relata* of logical consequence to be sentences. Other accounts are not excluded, as long as they involve sentences (e.g. sentence-context of utterance pairs).

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

Shapiro (1998) lists nine different intuitive characterizations of logical consequence of modal, epistemic and linguistic nature, among which are: “It is not possible for the members of  $\Gamma$  to be true and yet  $\Phi$  false” and “The truth of the members of  $\Gamma$  guarantees the truth of  $\Phi$  in virtue of the forms of the sentences (or propositions)” (Shapiro, 1998, p. 132).

I take the notion of *necessity*, as it is used above, to be a metaphysical notion, broadly understood, which may be explained in terms of possible worlds. I refer to possible worlds throughout this work, but do not assume a specific doctrine with respect to their nature. Specifically, I remain neutral with respect to realism or antirealism regarding possible worlds. My only assumption is that *necessity* can be understood as *truth in all possible worlds*.

However construed, possible worlds involve heavyweight questions regarding, e.g., the nature of existence, properties, objecthood, etc. But logical practice, it seems, attempts to satisfy *Necessity* without being completely steeped in metaphysics. It seems that logic is a discipline that tries to capture metaphysical necessities in a systematic way that avoids contentious metaphysical questions as much as possible. This is where formality enters the picture.

Note that *Formality* assumes that arguments have forms, and moreover, that each argument has a unique form. In this we follow a well-entrenched tradition in logic, though, I submit, not all characterizations of the formality of logic pertain to forms of sentences.<sup>3</sup> Yet, we are not told what *form* is. The answer to this question is far from obvious, and it requires further assumptions beyond the conditions *Necessity* and *Formality*. At this point, I do not want to commit myself as to what exactly constitutes form.

Considering the different approaches of the writers we are dealing with, we can come up with a minimal characterization of form, a common denominator for this debate. I will hereforth address the issue of formality through a certain *role* that is commonly attributed

---

<sup>3</sup>Presupposing that sentences have *unique* forms might be contested as well, but we will do so nonetheless to keep things simple.

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

to form, whatever form may be. The emphasis will be on this role, rather than a full and explicit definition or characterization of form. Although I try to make minimal assumptions about form and I presume that the role I attribute to it is quite appealing, I do not expect it to be in agreement with *all* prevalent views. I do contend that the idea I present is consonant with the approaches of Sher and Shapiro, the two main authors I discuss.

To the point: one feature which we might expect of form is that it will permit us to bypass metaphysical considerations to some extent. Whatever form is, it should relieve us from having to take sides in thorny metaphysical disputes or even from deciding which of them are coherent. So, for instance, we would like ‘Sarah likes cheese’ to follow logically from ‘Everyone likes cheese’, regardless of the metaphysical status of Sarah (Does she necessarily exist? Does she exist at all? Could she have had different parents than she actually has?) and that of cheese (is it a natural kind? Could it have grown on trees?). By this some necessarily true sentences, such as perhaps [*insert here your favourite metaphysical doctrine*], will not be logically valid. A step towards this desirable feature of form is to have forms of arguments be subject to mathematical handling, as they are in conventional formal systems of logic. A formal system might not capture all arguments satisfying *Necessity*, but only a restricted class of necessary arguments, which the system can detect on basis of their form. This is not to say that there are no other reasons to treat forms mathematically, but that treating forms of arguments as mathematical objects is in line with the desired feature.

Surely, formal systems for logical consequence may be involved with metaphysical questions to some extent. I do not deny any connection between metaphysics and logic – on the contrary: I take the condition of *Necessity* to be the primary condition on logical consequence. Yet we would like as much metaphysics as possible to be out of the way. It is more a matter of the *quantity* of metaphysical questions that form helps to reduce. We can require that the involvement of formal systems with metaphysics will be controlled and systematized at the outset, and the contentious metaphysical questions be reduced to a minimum. Moreover, we expect that even if some assumptions are endorsed at the outset

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

and embedded in the system, the practice of using the system to determine the validity of specific arguments be metaphysically innocuous: the verdict of the system on a given case where it applies will not be subject to metaphysical dispute.

I attribute to form a specific role, that of making it possible to capture *Necessity* without getting involved in metaphysics. Some might view the suggested role of formality as additional or as an outcome of *form*'s more basic features (e.g. its generality, topic-neutrality etc.)<sup>4</sup> Or, this role could also be taken as *the* basic feature of form. In any case, what is important for my current concerns is that a good account of form will be such where metaphysical questions are largely kept out of the way.

Despite their different treatments of formality, Sher and Shapiro seem to agree that logic should capture *Necessity* without immersing itself in metaphysics. As we shall see, this has significant implications for their philosophical approach to form. Neither Sher nor Shapiro specify the role of formality as I did, but this role can be inferred from their views nonetheless. We shall go into the details of their views in subsequent sections, but the outline of the interpretive inference is as follows: The metaphysical approach to semantics (see section 4.1) is problematic according to both Sher and Shapiro. The main reason it is problematic are the metaphysical questions it entails ((Sher, 1996, p. 661), (Shapiro, 1998, p. 149)). Incorporating an appropriate account of form (or, of logical terms, which constitute forms of sentences) will get rid of the problematic metaphysical questions ((Sher, 1996, p. 672), (Shapiro, 1998, p. 152)). So, by both Sher and Shapiro, in a proper account of logical consequence, form fulfills the role we attributed to it – of avoiding metaphysical questions.

---

<sup>4</sup>Sher is probably a good example. See section 4.3 for Sher's view of formality.

### 3.3 Models and Logical Consequence

Contemporary logical practice uses mathematical techniques to determine the validity of arguments on the basis of their form. Model theory allows a rigorous, precise treatment of validity, and moreover, provides a promise to keep logic out of a metaphysical mess. Let us assume that the language we are dealing with is of first-order, the details of which we leave open for the moment. A *model* is a pair  $\langle D, I \rangle$  such that  $D$ , the domain, is a non-empty set, and  $I$  is a standard interpretation function. According to Tarski’s model-theoretic definition of logical consequence, the sentence  $\varphi$  *follows logically* from the sentences of the set  $\Gamma$  (or: the argument  $\langle \Gamma, \varphi \rangle$  is *logically valid*) if and only if  $\varphi$  is true in each model where all sentences in  $\Gamma$  are true.<sup>5</sup>

In order for the model-theoretic definition to be adequate, the relation captured by it has to satisfy *Necessity* and *Formality*. To Tarski it was obvious that his definition “agrees quite well with common usage” (Tarski, 1936, p. 417), but he did not show that it indeed does so.<sup>6</sup>

The model-theoretic tradition has more to say about form than I have so far. It is a common assumption in the works of Tarski and his followers, and in contemporary logic in general, that the key factor determining the form of a sentence is its logical vocabulary. The question of form is thus usually dealt with through the investigation of logical terms.

---

<sup>5</sup>I am employing here “model” in the contemporary usage, not the original Tarskian one, as my concern is the philosophical underpinnings of the model theory we currently use. The differences will not concern us.

<sup>6</sup>Tarski only points out that the desirable features of consequence are outcomes of his definition. “In particular,” he writes, “it can be proved, on the basis of this definition, that every consequence of true sentences must be true, and also that the consequence relation which holds between given sentences is completely independent of the sense of the extra-logical constants which occur in these sentences” (*ibid*). It should be noted that it is not all that clear that Tarski adheres to *Necessity* as we formulated it, see (Tarski, 2002, pp. 25-29). Indeed, had Tarski provided a proof of the adequacy of his definition, it might have been clearer what he meant by the word “must”.

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

Tarski left the choice of logical terms as a (somewhat) free parameter in his definition of logical consequence (though he didn't think everything goes<sup>7</sup>). Later logicians tried to fix this parameter and define logical terms. Nonetheless, whatever the logical terms are taken to be, the model-theoretic definition is designed to satisfy *Formality*: truth-preservation over models is determined by the chosen logical vocabulary. Although I have previously not committed to what forms (or logical terms) are, I have attributed to them the role of helping us avoid metaphysical questions. I will consider this feature of form as a test for a choice of logical terms later on.

Apparently, for Tarski, the adequacy of the model-theoretic definition was obvious. But if we don't agree about the obviousness of these issues, we may articulate two philosophical gaps in the Tarskian account. Tarski's model-theoretic definition still leaves open how it:

- (1) captures necessity, and
- (2) makes it possible to avoid metaphysical questions (for some choice of logical terms).

Before we go on to examine ways by which these gaps may be filled, some issues that will not be dealt with should be set aside. Problematic metaphysical assumptions, one might claim, arise in model theory even before addressing the question of logical terms. For instance, the use of sets might be disputed. Models are set-theoretic entities, and as such, are subject to queries pertaining to the nature and existence of sets. Such worries may drive us to abandon set theory in favor of better foundations for model theory, or abandon models altogether in favour of, say, a proof-theoretic account. I will not attempt to deal here with metaphysical worries about sets, and just assume sets are metaphysically innocent.

There have been two prominent attitudes to the gaps in the Tarskian account. One is of viewing them as marking insurmountable difficulties in the Tarskian project ([Etchemendy, 1983, 1990](#)). The alternative, positive attitude, of trying to fill the philosophical gaps is adopted by contemporary adherents of model-theoretic consequence such as Sher and

---

<sup>7</sup>([Tarski, 1936](#), p. 418).

Shapiro.<sup>8</sup>

Now how would one go about showing that a formalism is philosophically adequate? A common ploy is to tie the formalism to the object of inquiry through *representation*. Models, apart from their role as mathematical objects, correspond to something in the non-mathematical world that has to do with logical consequence, and in this way we ensure *Necessity*. So, for instance, we could take each model to represent a possible world, and straightforwardly capture *Necessity*. In what follows we will see that this first attempt will not do, but a modification of it will be more successful.

Not all writers who take models to represent something state what such a relation requires. We can refer here to Shapiro, who provides us with a framework where the requirements on representation are made explicit. Shapiro presents the view of *logic-as-model* by which certain features of the mathematical apparatus we use in logic, the *representors*, represent features of logical consequence in natural language ((Shapiro, 1998), see also (Cook, 2002)). Other elements of the model are *artifacts*. It is important to note our double use of the word “model”. The “model” in “logic-as-model” does not refer to the models of model theory, but to the role of the whole formal system. I use the word “model” in both senses, assuming that context can disambiguate.<sup>9</sup> By the view of *logic-as-model*, a formal system as a whole serves as a *mathematical model* for logical consequence. A model in general simplifies and idealizes the modeled phenomenon, aiming at an easier to handle, yet adequate, account. What it takes for a model to be adequate is not always clear-cut,

---

<sup>8</sup>To be precise, Etchemendy thinks that the conceptual adequacy of model theory can be salvaged, but as he stresses, it would be in a way very foreign to Tarski’s so called reductive analysis. In fact, Etchemendy’s approach to model theory is quite close to that of Shapiro, but he fiercely criticizes the attribution of such an approach to Tarski, see (Etchemendy, 2006).

<sup>9</sup>Obviously, there is a connection between the two uses of “model”. But when referring to the models of model theory we use “model” as a technical, mathematical, notion, where the aspect of models of model theory modeling something will be attributed to them only in virtue of being part of a bigger model, which consists of the whole formal system.

## CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

and it depends on the purpose of the investigation. Shapiro explains that “typically, there is no question of ‘getting it exactly right’ ” and that “there should be a balance between simplicity and closeness of fit” (Shapiro, 1998, p. 138). In Shapiro’s view, logical consequence is a cluster notion that involves varying intuitions. He claims that model-theoretic consequence can serve as a good mathematical model of the notion of logical consequence based on modal and semantic intuitions, which we articulated in the conditions of *Necessity* and *Formality* (Shapiro, 1998, p. 148). The models of model theory (“models” in the other sense) are merely items in the formal system that are nominees to be representors or artifacts in the big model.

Sher too connects the formal system to the concept of logical consequence through representation, though she does not employ an elaborate account such as Shapiro’s logic-as-model. Sher seems to demand, however, a closer match between representors and what they represent, and this, we will see, poses a disadvantage to her account.

### 3.4 What Models Represent

Taking on the logic-as-model approach, it is natural to assume that the models of model theory are not mere *artifacts* of the system but are *representors*. I will work under that assumption, and the guiding question will be: what do models represent? The following subsections present four answers to this question. I first discuss in brief two unsatisfactory initial attempts at a solution, inspired by Etchemendy (1990), though I will follow Sher’s presentation and terminology. The shortcomings of the first proposed answers lead to alternative solutions by Sher and Shapiro, which will be our main focus.

#### 3.4.1 Metaphysical Semantics

A way of understanding the condition of *Necessity* is as a requirement of truth preservation in all possible worlds, while model-theoretic consequence preserves truth in all models



### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

– which suggests correlating models to possible worlds. In *metaphysical semantics* (or “representational semantics” in Etchemendy’s terms), models represent possible worlds, or possible configurations of the world (see (Etchemendy, 1990, ch. 2)).

The metaphysical approach does not fit common practice very well, and would entail a revision of standard model theory (see (Sher, 1996, p. 659), (Shapiro, 1998, p. 143)). It is therefore most appropriate to view metaphysical semantics as not merely a philosophical perspective on model theory, but as an approach on how the apparatus of model theory should be adjusted and employed, regardless of current practice. Metaphysical semantics requires a fully interpreted language, so that evaluating sentences at possible worlds would be possible. In this approach, a sentence is true in a model if and only if the model represents a possible world where the sentence is true (or the proposition expressed by the sentence holds).<sup>10</sup> A sentence such as ‘The board is red all over and green all over’ (or its formalization) would be true in some model only if it is metaphysically possible for the board to be both red and green all over. That is, this sentence would be rendered logically false.

It is easy to see that metaphysical semantics, if viable, satisfies *Necessity*. To the extent that all possible worlds can be represented by models in the system, the model-theoretic consequence relation would be necessarily truth-preserving. Metaphysical semantics simply equates logical consequence with *necessary* consequence. In this approach, logically valid arguments just are those that are necessarily truth-preserving. This would not be agreeable to those who think that *Formality* is a substantive condition on logical consequence (whatever forms may be).

From the point of view of our present concerns: if we can use the model-theoretic apparatus for metaphysical semantics, albeit with some revision, we can take care of the gap

---

<sup>10</sup>We assume here, together with Sher, Shapiro and Etchemendy that there is a natural, straightforward way of evaluating sentences in possible worlds or configurations of the world. For a critical discussion of this assumption see (Blanchette, 2000).

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

posed in (1) above, i.e. capture necessity. However, (2) – avoiding metaphysical questions – is completely ignored in this approach. Sher lays out the main disadvantage of metaphysical semantics:

[M]etaphysical semantics requires solutions to the most obscure and thorny questions of general metaphysics. [...] We come upon recalcitrant questions of identity, essential properties, moral and rational agency, meaning, etc. that have baffled philosophers for years. (Sher, 1996, pp. 660-661)

Sher’s concern here is that of avoiding metaphysical questions. Whatever forms of sentences might be in this approach, they do not fulfill the role we have attributed to them: metaphysical semantics (as its name already discloses) does not save us from a single metaphysical question. In order to construct our semantics, we need to know the modal status of every sentence in our language (see also (Etchemendy, 1990, p. 25)).

These considerations suffice for us to move on and examine other possible approaches to semantics. Clearly, metaphysical semantics is not adequate at filling the philosophical gaps of the model-theoretic definition of consequence.

#### 3.4.2 Linguistic Semantics

Another option for the model-theorist is to follow the *linguistic semantics* approach, and view the variations in models as representing variations in *interpretations* of the language.<sup>11</sup> This, again, is an approach that uses the apparatus of models with some modifications to contemporary standard practice – and so cannot be viewed as a perspective on ordinary model theory.<sup>12</sup> The domain of discourse in this view is held fixed across models. Linguistic

---

<sup>11</sup>The idea of linguistic semantics is inspired by Etchemendy’s “interpretational semantics”. But by contrast, in interpretational semantics models do not represent anything, they *are* interpretations. Etchemendy attributes interpretational semantics to Tarski. The difference between *linguistic* and *interpretational* semantics is not crucial for our purposes and specifically, we are not concerned with Tarski exegesis.

<sup>12</sup>See (Sher, 1996, p. 666) and (Shapiro, 1998, pp. 143, 149).

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

semantics frees itself of much of the metaphysical burden posed by metaphysical semantics, and is concerned only with the world as it *is*, not as it *could have been*, and thus each model represents the actual world. The relation between language and the world, the different ways language can be interpreted – that is what makes the difference between models.

We can use an example by Shapiro to display the difference between linguistic and metaphysical semantics. Consider the sentence ‘Snow is black’ (or a formal rendering of it). In the metaphysical approach to semantics, a model in which this sentence is true represents a possible world where the color of snow is black. In the linguistic approach, a model in which the sentence is true represents an interpretation of the language in which the sentence *means* something true, e.g. one where ‘black’ denotes *white* (Shapiro, 1998, p. 142).

In linguistic semantics, all re-interpretations of the language must respect *form*. On the assumption that logical terms determine the forms of sentences,<sup>13</sup> those terms keep their meanings across models – they are “held fixed”. So what varies from model to model is just the interpretation of the *nonlogical* terms. In the example above, ‘black’ can be re-interpreted as it was only if it is considered as a nonlogical term. The linguistic perspective leans heavily on the distinction between logical and nonlogical terms. As a result, it is much more faithful to *Formality*.

Linguistic semantics does not suit standard model theory, since the latter does not keep the domain fixed. But suppose we made the proper adjustments to our model theory so it will fit the linguistic perspective. That is, fix the domain to be a set that represents the actual world. Have we made any progress with respect to the gaps we mentioned in (1) and (2)? It seems that with linguistic semantics the tables have turned: we are now doing better at avoiding metaphysical questions. We need be only concerned with metaphysical questions pertaining to one possible world, the actual one, rather than all possible worlds. By contrast, satisfying *Necessity* appears to be the weak link. The idea motivating linguistic

---

<sup>13</sup>To be precise, what determines the form of a sentence is not only logical terms and their arrangement in the sentence, but also the grammatical categories of nonlogical terms and their iterations in the sentence.

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

semantics might be that modality could be reduced to semantic variability. Perhaps, with the right choice of logical terms, dealing only with the actual world would suffice – variations in the world would be accounted for through variations in the language. This hope, however, cannot be sustained.

We will not go over all the possible choices of logical terms, but will note that linguistic semantics is not adequate with the standard first order selection of logical terms. Consider the following sentence  $\varphi$ :  $\forall x \exists y Rxy \wedge \forall x \forall y \forall z ((Rxy \wedge Ryz) \rightarrow Rxz) \wedge \forall x \neg Rxx$ , which states that the binary relation  $R$  is serial, transitive, and irreflexive (Etchemendy, 1990, p. 118). Such a relation is possible only in an infinite domain. Now assume that the universe consists of only finitely many objects, so the fixed domain for our semantics is finite. In that case,  $\neg\varphi$  would be true in all models, that is, it would turn out to be logically true ( $R$  is the only nonlogical term in the sentence). This would be in violation of *Necessity*, if we countenance infinite possible worlds. We could then argue that the world is infinite. That would change the logical status of  $\neg\varphi$  to not logically true (and not logically false). In this case,  $\neg\varphi$  would not be a counter-example to *Necessity*. But then, arguably, we are letting our logic rely on a substantial metaphysical assumption regarding the size of the world, in contrast to the spirit of *Formality*. Even though we might not be able to avoid all metaphysical questions, this one might involve a price too heavy to pay.<sup>14</sup>

The prospects of finding a choice of logical terms which would both satisfy *Necessity* and not involve weighty metaphysical assumptions are bleak. Narrowing down to just the actual world does not free us from metaphysical investigations. Some questions about what *actually is* are already heavily loaded and intricate.<sup>15</sup> Thus, linguistic semantics does not

---

<sup>14</sup>Etchemendy proposes a modification that compensates for the fixed domain. The quantifiers are analyzed as having two components, so that, e.g.,  $\exists$  is understood as *some-thing*, where *some* is fixed and *thing* varies over subsets of the domain. Note that as subsets of the domain never have cardinality greater than that of the domain, even with this modification the logical truth of  $\neg\varphi$  depends on finitude of the world, see (Etchemendy, 1990, pp. 112-120).

<sup>15</sup>Indeed, metaphysical investigations are a way to learn things about the actual world, sometimes more

deliver the adequacy of model-theoretic consequence. Nevertheless, as we learn from Sher and Shapiro, metaphysical and linguistic semantics do not exhaust the possible approaches model-theorists can employ.

### 3.4.3 The Formal-Structural View

A natural solution to the inadequacy of each of Etchemendy’s proposed semantic views is to combine the two so that the features of each of them compensates for the disadvantages of either one or the other. Such is Shapiro’s *blended approach*, according to which models represent possible worlds under interpretations of the nonlogical terms of the language. *Necessity* and *Formality* are captured in this approach, as they are captured separately in representational and interpretational semantics respectively.<sup>16</sup> The blended approach is discussed extensively in the next subsection.

Sher seems to be aware of the possibility of combining metaphysical and linguistic semantics, but she chooses a different route – her *formal-structural view*. For Sher, “a (logical) model represents a formally possible structure of objects relative to the primitive terms of a given language” (Sher, 1996, p. 675). These structures consist of objects and correspond in some way to possible states of affairs, but they are emphatically not possible worlds. Indeed, Sher is an adamant critic of the view of models as representing possible worlds, as was previously pointed out in the discussion of metaphysical semantics. She argues that any position that takes models as representing possible worlds entails formidable difficulties; it imports metaphysical conundrums into logical practice, and requires taking into account an unreasonable amount of information for any reasonably rich language. The exceedingly complex metaphysical issues that arise in the description of possible worlds

---

feasible than empirical methods.

<sup>16</sup>This obvious solution is absent from Etchemendy’s presentation of the two approaches to semantics in (Etchemendy, 1990). His own approach to semantics, which he describes in (Etchemendy, 2006) as a version of representational semantics, is, however, very similar to Shapiro’s.

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

cannot be expected to be solved in the formal system, and therefore cannot serve as a basis for model-theoretic semantics. This worry is not limited to metaphysical semantics, in which logical consequence is equated with necessary consequence, and in which each model is viewed as representing a possible world; Sher's critique is also aimed at views involving possible worlds that take the category of logical consequence as merely derived from the category of necessary consequences.<sup>17</sup> We see that Sher requires of formality to fulfill the role we attributed to it at the start: drawing the set of formal and necessary consequences from the set of necessary consequences will not give a good account of logical consequence if all the metaphysical questions remain.

Yet how can Sher present models as representing or corresponding somehow to possible states of affairs without viewing them as representing possible worlds (or possible ways the world might be, or any other equivalent metaphysical notion)? In the course of the present subsection I suggest that Sher's view of models as formally possible structures is unsatisfactory, and that her view inevitably collapses to the blended approach, which does incorporate possible worlds.

According to Sher, and in line with the presuppositions of this work, logical consequences are the formal and necessary consequences, and in Sher's words "formal necessity is a particular case of both necessity and formality" (Sher, 1996, p. 668). The concept of *formality* is a key notion in Sher's approach, and she tackles it through the definition of logical terms.

For Sher, "Logical terms are formal in the sense of denoting properties and relations that are, roughly, intuitively structural or mathematical" (Sher, 1996, p. 668). Sher is a follower of the Tarskian tradition, according to which the logical terminology that appears

---

<sup>17</sup>Sher implies that a semantic approach which takes logical consequence as a subset of necessary consequence has two options: either it shares the disadvantages of metaphysical semantics (Sher, 1996, p. 659), or it is simply useless, "tantamount to giving an account of the distinctive nature of *logical* semantics" (Sher, 1996, p. 661).

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

in a sentence determines its form. Like Tarski, she holds that as a result of *Formality*, logical consequence is independent of our ability to distinguish between objects in the world. In Tarski's words:

Moreover, since we are concerned here with the concept of logical, i.e. *formal*, consequence, and thus with a relation which is to be uniquely determined by the form of the sentences between which it holds, this relation cannot be influenced in any way by empirical knowledge, and in particular by knowledge of the objects to which the sentence X or the sentences of the class K refer. The consequence relation cannot be affected by replacing the designations of the objects referred to in these sentences by the designations of any other objects. (Tarski, 1936, pp. 414-415)

Sher (Sher, 1991, 1996) proposes a mathematical definition of logical terms. Formally, Sher's system of *universal logic* extends standard classical predicate logic. Sher concentrates on variations of first order logic; in her system, each logical term is a truth-functional connective, or an  $n$ -place predicate or functor of level 1 or 2. The general criterion for formality in Sher's approach (also known as the Tarski-Sher Approach) is:

(F) A term is formal iff it is invariant under isomorphic structures. ((Sher, 1996, p. 677), see also (Tarski, 1986))

The terminology satisfying (F) includes standard logical terminology, as well as non-standard logical terms, such as *finite* and *most*. Truth functional connectives can be incorporated in different ways,<sup>18</sup> but to keep things simple, Sher just defines them to be logical. Thus, we have *Sher's criterion for logical terms*:

---

<sup>18</sup>Truth functional connectives are invariant under isomorphic propositional structures, and can fit the schema of Sher's definition of logicity. Alternatively, the connectives can be viewed as generalized quantifiers (see (Sher, 2008)).

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

$C$  is a logical term (in universal first order logic) if it is a truth-functional connective, or a predicate or functional expression of level 1 or 2 that satisfies (F) (see (Sher, 1996, p. 679)).

It has been less noted that models also have a central role in Sher’s approach to formality. Models in Sher’s account are representors – they represent *formally possible structures*. Now let us take a step back. The basic conditions at the basis of our discussion are *Necessity* and *Formality*. We have agreed (i.e. all writers referred to here agree) that possible worlds are an admissible way of speaking of necessity. At this point we are also assuming (the widely held assumption, no doubt) that logical terms determine the form of sentences. Let us look at whether the notion of *formally possible structure* is a reasonable addition to our conceptual apparatus, and whether it can ground the adequacy of the model-theoretic definition.

A straightforward way to obtain all logical consequences is to start out from the class of all necessary consequences (where the conclusion follows necessarily from the premises). Logical consequences are a subset of this class: they are the necessary consequences that are also “formal”. In this method one might rely on a possible-worlds-oriented semantics, as in metaphysical semantics, and in this way obtain all necessary consequences. From those, all logical consequences can be obtained by using an additional criterion, such as: a logical consequence is a necessary consequence that is also necessary under every re-interpretation of the nonlogical terms. This is, as noted before, Shapiro’s approach, that will be discussed in the next subsection.

Sher’s approach is, at least ostensibly, completely different. She contends that the straightforward approach entails all the problems of incorporating possible worlds into the logical system. However, she suggests that the collection of arguments that satisfy necessity can be narrowed down in a different manner – by looking at a wider notion of possibility. The range of possible worlds gives a certain collection of necessary consequences, the ones



### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

for which there is no possible world where the premises are true and the conclusion false. A wider range of states or possibilities would bring about a narrower collection of arguments, which would be included in the one obtained by the original range.

Sher thus extends the standard concept of *possibility*. She avoids using *possibility* as that which accompanies the concept of *necessity* and is ordinarily explicated through possible worlds, dealing instead with what she terms *formal possibility*, which accompanies the concept of *formal necessity*, and can include cases that are not metaphysically possible. Thus, for example, there is a formal possibility in which the statement ‘The board is red all over and green all over’ is true. Each formal possibility is represented by a model:

[T]he totality of models reflects “possibilities” that in general metaphysics might be ruled out by nonformal considerations. That is to say, the notion of *possibility* underlying the choice of models is wider than in metaphysics proper. (Sher, 1991, p. 138)

In this manner Sher avoids the metaphysical discussion of possible worlds; her only obligation is to characterize formally possible structures. Formally possible structures are represented by models, which are set-theoretic constructs. In this way, Sher reduces formal possibility to mathematical objects:

The reductive approach to modal notions is based on the idea that imprecise intuitive notions can be reduced to clear yet adequately strong precise notions... On my conception, formal possibility is reduced to mathematical existence and formal necessity to mathematical generality: “It is formally possible that  $\Phi$ ” to “There exists at least one mathematical (set theoretical) structure  $\mathfrak{S}$  such that  $\Phi$  holds in  $\mathfrak{S}$ ” and “It is formally necessary that  $\Phi$ ” to “For all mathematical (set theoretical) structures  $\mathfrak{S}$ ,  $\Phi$  holds in  $\mathfrak{S}$ .” (Sher, 1996, p. 682)

Let us take a closer look at Sher’s manner of reasoning. The concept of logical con-

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

sequence revolves on the concept of necessity and on the concept of formality. Logical consequences are the formally necessary consequences; there is therefore no need to deal with the general concept of necessity, but only with the formally necessary – what is formally necessary is, obviously, both formal and necessary. And one might think that in the same manner in which the formally necessary can be established, so can the dual concept of the formally possible. But here we meet a problem. Whereas the extension of the formally necessary is included in that of the ordinary concept of necessity, the extension of the formally possible *extends* that of the ordinary concept of possibility, and it not clear how it is to be attained. It's obvious that to get the formally necessary we intersect the formal with the necessary. It is less clear how to extend the concept of possibility: Sher does not tell us how to come up with the formal possibilities which are not metaphysical possibilities.

By employing the concept of formal possibility, Sher seemingly gains the advantage over views that involve possible worlds, incorporating necessity into her system without the “metaphysical baggage”. But when adopting the new concepts of formal necessity and formal possibility, she abandons the familiar ground of accepted concepts. *Formal possibility*, however, is not the result of the intersection of two pre-given concepts, but of the expansion of one of them, and cannot be straightforwardly derived from what is *formal* and what is *possible*. That is to say, the concept of formal possibility is a completely new modal concept which Sher employs as her proposed theoretical basis for model-theoretic semantics. Sher does not give a substantive definition of *formal possibility* – that which models represent in her approach. Nor does Sher adopt it as a primitive concept. As we shall see, an attempt to achieve a clear understanding of this concept leads to the conclusion that *formal possibility* can best be understood as *interpretation of the nonlogical terms in a possible world* (or *state of affairs*), despite Sher's obvious wish to leave out possible worlds.

Consider the example sentence ‘The board is red all over and green all over’. Under ordinary modal terminology this sentence is “impossible”: it is false in all possible worlds. But this sentence would be, by Sher's analysis, formally possible. The meaning of this is

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

that there is a formally possible structure for the language (in this case, natural language or a fragment thereof) in which this sentence is true, and this structure is represented by a model, in which the corresponding formalized sentence is true, for:

Intuitively, a (logical) model represents a formally possible structure of objects relative to the primitive terms of a given language. (Sher, 1996, p. 675)

But what is a “formally possible structure relative to the primitive terms of a given language”?

According to Sher, models represent something that has to do with possible states of affairs:

Tarski’s semantics... [uses] a semantic apparatus which allows us to represent the relationship between language and the world in a way that distinguishes formal and necessary features of reality. The main semantic tool is the model, whose role is to represent possible states of affairs relative to a given language. (Sher, 1991, p. 138)

[W]e would like to know how models represent all intuitive possibilities with respect to a given language: what features of possible states of affairs correlate with what features of models, how differences between models suffice to represent all relevant differences between states of affairs. (Sher, 1996, p. 657)

If we combine this with Sher’s stance that nonlogical terms are re-interpreted in every model, we may venture to summarize her position thus: models represent constructs that correlate to possible states of affairs, and nonlogical terms are re-interpreted in them. My claim is that is simply another way for saying that a formally possible structure is, in fact, an interpretation of the nonlogical terms in a possible world (or state of affairs).<sup>19</sup> The

---

<sup>19</sup>The difference between possible worlds and states of affairs (the former committing to maximality not

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

sentence ‘The board is red all over and green all over’ is true in, for example, a formally possible structure where the domain is  $\{a\}$ , ‘the board’ denotes  $a$ , and ‘red all over’ and ‘green all over’ both denote  $\{a\}$ . And this structure corresponds to an interpretation of the nonlogical terms in a possible world where there is just one object, and in which the extensions of ‘red all over’ and ‘green all over’ consist of the only object in the world.

Sher attempts to use formally possible constructions to circumvent possible worlds, or any metaphysical import for that matter. Sher’s circumvention is only superficial: the most plausible way to understand Sher’s use of formally possible constructions is as possible worlds under interpretations of the nonlogical terminology. Thus, the alternative Sher offers to metaphysical and to linguistic semantics is ultimately a combination of both. However, a view of models as representing possible worlds under interpretations does not necessarily entail the metaphysical difficulties Sher attributes to it. This point will be elaborated in the next subsection.

At the current stage my claim is the following: when one tries to understand the new modal concept Sher employs, and endow it with substance in a way that complies with her remarks, this concept collapses to the ordinary concepts of necessity and formality that are ordinarily explicated through possible worlds and interpretations respectively. My criticism still leaves one loose end: it seems that Sher succeeds in establishing a theoretical basis to logic, while steering clear of the difficulties arising from taking models to represent possible worlds. As a reply to my criticism, Sher might claim that if her account avoids metaphysical issues, there is a marked difference between it and any account in which possible worlds are indeed incorporated. Thus, the introduction of formal possibilities might have some beneficial feature I have overlooked. In the next subsection I show that despite Sher’s contention, there is no need to throw away possible worlds in order to steer clear of metaphysical questions.

---

assumed by the latter) should not concern us here; most of the metaphysical questions alluded to section 4.1 with respect to possible worlds will re-arise when using possible states of affairs.

### 3.4.4 The Blended Approach

Let us now look more closely into Shapiro’s suggestion to “blend” metaphysical and linguistic semantics. Shapiro contends that logical consequence is a cluster-notion which involves modal, semantic and epistemic intuitions. *Necessity* and *Formality* are good representatives of the modal and semantic intuitions, and together they are a basis for a *conglomeration* notion of consequence, hereafter referred to as *Cong.*:

*Cong.*:  $\Phi$  is a logical consequence of  $\Gamma$  if  $\Phi$  holds in all possibilities under every interpretation of the nonlogical terminology in which  $\Gamma$  holds. (Shapiro, 1998, p. 148)

Model-theoretic consequence, according to Shapiro, can be used as a good mathematical model of *Cong.* We simply take models to represent possible worlds under interpretations of the nonlogical terms. This blend of metaphysical and linguistic semantics is not as artificial as it may seem. Shapiro advises us to be sceptical of the metaphysical-linguistic dichotomy in the first place. “Most philosophers”, he says, “do not doubt that there is a difference between the role of language and the role of the world in determining the truth-values of sentences, but many hold that this distinction is not at all sharp” (Shapiro, 1998, p. 147). Take a model  $M$  where ‘Snow is black’ is true, and assume that both ‘snow’ and ‘black’ are nonlogical. In the blended approach it is indistinguishable whether  $M$  represents a way the world could be in which snow is black, or an interpretation of the language by which ‘Snow is black’ turns out to be true. It is sufficient that we know all possibilities are accounted for, and in general that we have a good model of logical consequence.

Here, though, Sher’s concerns reappear. Obviously, both *Necessity* and *Formality* are satisfied in the blended approach. What is less obvious is how the gap previously mentioned as (2) is addressed: It seems that we need to answer a multitude of metaphysical questions in order to represent all possible worlds, before we add the additional dimension of varying the language and blur the *truth in virtue of the world*/*truth in virtue of language* distinction.

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

But this contention is wrong. There is no need to take two separate steps, one to account for all metaphysical possibilities and another for possible interpretations. We can make use here of the logic-as-model approach: we don't need to account for all possible worlds, as long as *Necessity* is made sure to be satisfied. We need not be worried whether it is possible for snow to be black. Varying the interpretations of 'snow' and 'black' with a minimal variation of the domain will account for all possibilities with respect to 'Snow is black'.

The blended approach resembles linguistic semantics when re-interpretations step in instead of possible worlds. The difference is that the blended approach does not abjure possible worlds: it does not claim that modality can be completely reduced to semantic variation.

Let us make this idea more precise. We need our models to be such that varying interpretations can take care of all possibilities while avoiding as much metaphysics as possible. Learning from the failure of linguistic semantics, we let the size of the domain vary, taking account of all the possible sizes the world could have. The question of the sizes of possible worlds involves us with some metaphysics, but as we shall see, this is the only metaphysical question we have to deal with.

Shapiro suggests the following property for “any model theory worthy of the name” (Shapiro, 1998, p. 152):

**The isomorphism property.** A language  $L$  (in the formalism) is said to have the isomorphism property iff for every formula  $\varphi$  and models  $M$  and  $M'$  for  $L$  with corresponding assignments  $s$  and  $s'$  such that  $\langle M, s \rangle$  and  $\langle M', s' \rangle$  are isomorphic with respect to the non-logical terms in  $\varphi$ ,  $\langle M, s \rangle \models \varphi$  if and only if  $\langle M', s' \rangle \models \varphi$  (Shapiro, 1998, p. 151).<sup>20</sup>

---

<sup>20</sup>In his (Shapiro, 1998), Shapiro does not mention assignments, but it is natural to add them in, since a formula is evaluated by a model only under an assignment. Alternatively, we may restrict formulas to sentences, as Shapiro does in (Shapiro, 2005). This difference doesn't have significant implications for our purposes.

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

As Shapiro explains, “The isomorphism property is a manifestation of the intuition that logical truth and logical consequence should be a matter of ‘form’, to the extent that isomorphism preserves ‘formal’ features of various models (whatever those are)” (Shapiro, 1998, pp. 151-152).

The isomorphism property can be viewed as a constraint on logical terms. In fact, as we’ll see in the next section, it is equivalent to Sher’s criterion for logical terms. Shapiro, however, takes the isomorphism property to be a necessary, though perhaps not sufficient condition on formal languages *vis-à-vis* their logical terms. How does this constraint help us in addressing (2)?

Shapiro acknowledges that logic should avoid metaphysical issues, at least to some extent (see (Shapiro, 1998, p. 149)). The isomorphism property, Shapiro shows, has the following corollary:

**Corollary.** Let  $L$  be a first order language. Let  $D$  and  $D'$  be two domains of the same cardinality, and assume that the isomorphism property holds for  $L$ . Then a sentence  $\varphi$  is true under every interpretation of the nonlogical vocabulary of  $L$  in  $D$  if and only if it is true under every interpretation of the nonlogical vocabulary of  $L$  in  $D'$ . (See (Shapiro, 1998, p. 152))

The corollary teaches us that given the isomorphism property, it is enough to take one domain of each size and all the model-theoretic interpretations on it to account for all possibilities model theory can account for. So we do not have to represent each and every possible world by a separate model. Each model will represent all possible worlds of its size, under some interpretation of the nonlogical vocabulary. *Necessity* will still be satisfied, and we arrive at a good position with respect to metaphysical questions: the only information we need regarding possible worlds is the class of cardinalities of all possible worlds, that is,

## CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

we just need to know the sizes possible worlds may have.

Standard first order logic achieves even a better result. From the downward Löwenheim-Skolem theorem we can further conclude that a countable domain suffices to account for all infinite possible worlds. (Each finite possible world, if there are any, must still be accounted for by a finite model of the same cardinality). This may provide reason to restrict the class of logical terms even further, and require that terms would be admitted as logical only if the Löwenheim-Skolem theorems hold for the resultant logic. This, however, is not a mild restriction: it would rule out a great many terms that the isomorphism property allows. Generalized quantifiers such as *there exist uncountably many* and many other terms that have to do with cardinality do not agree with the Löwenheim-Skolem theorems (not to mention second-order logic).<sup>21</sup>

Even when the isomorphism property is the only constraint on logical constants, the blended approach succeeds in both capturing the required kind of necessity and in avoiding *most* of the complex metaphysical issues it involves. This indeed captures the gist of the role we set for formality: the selected logical terminology makes it possible to ground the necessary truth-preservation of a restricted class of arguments using (mostly) formal means, without having to deal with an overwhelming burden of metaphysical questions.

### 3.5 Tying the threads

A comparison of Sher and Shapiro's semantic approaches reveals that Shapiro's is more satisfactory at filling the philosophical gaps in Tarski's account of consequence. It is more straightforward: *Necessity* is captured without resorting to new, unexplained modal notions. One could of course take *formal possibility* and *formal necessity* as the primitive notions

---

<sup>21</sup>If we further require countable compactness, the resultant logic will not be stronger than first order logic by Lindström's theorem, and the isomorphism property will be almost devoid of content, as will Sher's criterion.



### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

underlying logical consequence, but Sher starts with the plain notions of metaphysical possibility and necessity as we do. She then employs the idea of a *formally possible structure* whose role is that of avoiding the metaphysics involved in incorporating possible worlds. As long as this notion remains primitive or unexplained, it is not clear that it captures *Necessity*. We have seen that upon scrutiny, this notion collapses to that of *possible world under interpretation of the nonlogical terms*.

Shapiro and Sher do not share exactly the same conceptual apparatus, as they differ in their treatment of the notion of *formality*. A thorough analysis of *formality* leads Sher to her invariance criterion for logical terms. Shapiro, on the other hand, proposes the isomorphism property as a necessary condition on languages. Interestingly, though, Sher and Shapiro's conceptions of formality are not all that different, at least on the technical side. When we bring their criteria to a common framework, namely first order languages (by Sher's restriction), they turn out to be equivalent.

**Equivalence result.** Let  $L$  be a first order language with logical terms of level 1 or 2 (truth-functional connectives, predicates and functional expressions), and countably many nonlogical predicates from each arity. We assume also that the model theory is a standard classical one. Then the isomorphism property holds for  $L$  iff the predicates and functional expressions which are logical terms in  $L$  are invariant under isomorphisms.<sup>22,23</sup>

As shown by the equivalence result, the choice between Sher's criterion and Shapiro's isomorphism property becomes a practical matter depending on the specific task at hand. The isomorphism property has the advantage of conciseness and generality. On the other hand, if we have a language  $L$  for which we already know that the logical terms are well behaved in

---

<sup>22</sup>Note that our use of the term *logical term* is neutral and does not presuppose any criterion for admissibility.

<sup>23</sup>See the appendix for an outline of the proof, which appears in full in (Sagi, 2011).

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

Sher and Shapiro's terms, Sher's criterion makes it easier to introduce an additional logical term to  $L$ .

Recall that the corollary to the isomorphism property shows that metaphysical questions can be reduced in systems satisfying the isomorphism property. By virtue of the equivalence result, the corollary to the isomorphism property also follows from Sher's criterion for logical constants. Hence the evasion of metaphysical questions can be assumed in Sher's system, even when models are taken to represent possible worlds. Not every aspect of a possible world needs to be represented while the desired extension of logical consequence can still be achieved. Sher doesn't have to concern herself with the contentious notion of *formal structure* or *formal possibility*.

## 3.6 Conclusion

The formality of logic is commonly described as independence from a certain kind of information or content. For Sher, it is independence from the identity of particular objects. Shapiro aims to have formal languages that do not involve us in intricate questions on modality. I haven't offered in this chapter a precise type of content which logic should disregard. I used the term "metaphysical questions" to refer, quite broadly, to the type of questions logic tries to avoid. The attitude I aimed to capture, shared by prominent writers in the area, is one by which logical consequence encompasses all possibilities (satisfies *Necessity*) without getting entangled in all that these possibilities involve. We would like valid arguments to preserve truth no matter what, in any case, but we are very far from clear on what all cases are. We then seek a mathematical apparatus that captures valid arguments without relying on contested issues.

There is no consensus on what the formality of logic amounts to. This is reflected in recent debates on the issue of logical terms. Sher and Shapiro approach the problem from

quite different angles, but I believe that their accounts are both motivated to some extent by the attitude I posed. Both of them propose formal systems that are purported to capture *Necessity* without getting entangled in metaphysics. Shapiro is most successful in showing how this can be done – through the *logic-as-model* perspective combined with viewing models as representing possible worlds under interpretations of the nonlogical vocabulary. The latter principle ensures *Necessity*, the former relieves from the need to represent all possible worlds.

Sher and Shapiro (and many others) deal with formality through forms of sentences, and more specifically, through logical terms. Logical terms have gained such centrality in the philosophy of logic, so that the question of the formality of logic often turns to one of the formality of some *terms*. Adopting the attitude I delineated here, however, does not commit one to there being a *category* of formal terms. Some characterizations of logical terms are simply more beneficial than others in keeping logic formal and away from problematic metaphysics. Sher’s invariance criterion and Shapiro’s isomorphism property suit our general purpose.

To be sure, I have not claimed that metaphysical questions can or should be avoided *completely*. What seems to fit the proposed picture of logic is a more pragmatic approach, by which we try to balance the poles of *Necessity* and *Formality*. When devising formal systems, we seek a mathematical apparatus which will compensate for our less than satisfactory grasp of the notion of *necessity* involved in logic. This aim can be achieved in various ways, and does not require a search for the one “correct” extension of logical terms or consequence in general.

### 3.7 Appendix

**Equivalence result.** Let  $L$  be a first order language with logical terms of level 1 or 2 (truth-functional connectives, predicates and functional expressions), and countably many

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

nonlogical predicates from each arity. We assume also that the model theory is a standard classical one. Then the isomorphism property holds for  $L$ , with the class of models for  $L$ , if and only if the predicates and functional expressions which are logical terms in  $L$  are invariant under isomorphisms.

*Proof Outline.* The left to right direction (from the isomorphism property to Sher's criterion) is proved by constructing, for each logical term  $C$ , a formula  $\phi$  where  $C$  is the only logical term, and then using the isomorphism property on  $\phi$ . We show this only for the simple case of first-level predicates.

Let  $C$  be an  $n$ -place first-level predicate for some  $n \geq 1$ , and assume that  $C$  is a logical term in  $L$ . We need to show that it satisfies Sher's criterion, i.e. that it is invariant under isomorphic structures. Sher assigns to each logical term  $C$  a function  $f_C$  such that for each model  $M$ ,  $f_C(M)$  is  $C$ 's extension in  $M$ . Let  $M$  and  $M'$  be models with domains  $A$  and  $A'$ , and  $\langle b_1, \dots, b_n \rangle \in A^n$ ,  $\langle b'_1, \dots, b'_n \rangle \in A'^n$  such that the structures  $\langle A, \langle b_1, \dots, b_n \rangle \rangle$  and  $\langle A', \langle b'_1, \dots, b'_n \rangle \rangle$  are isomorphic. We need to show that  $\langle b_1, \dots, b_n \rangle \in f_C(M)$  iff  $\langle b'_1, \dots, b'_n \rangle \in f_C(M')$ . We now define assignments  $s$  and  $s'$  for  $M$  and  $M'$  respectively, thus:

$$s(x_i) = \begin{cases} b_i & \text{if } 1 \leq i \leq n \\ b_n & \text{if } i > n \end{cases} \quad s'(x_i) = \begin{cases} b'_i & \text{if } 1 \leq i \leq n \\ b'_n & \text{if } i > n \end{cases}$$

Note that the structures  $\langle A, \langle b_1, \dots, b_n \rangle, s \rangle$  and  $\langle A', \langle b'_1, \dots, b'_n \rangle, s' \rangle$  are isomorphic. Now consider the formula  $\varphi = C(x_1, \dots, x_n)$ . From the previous claim we get that  $\langle M, s \rangle$  and  $\langle M', s' \rangle$  are isomorphic with respect to the nonlogical terms in  $\varphi$  (the nonlogical terms in  $\varphi$  include only variables<sup>24</sup>). So by the assumption that the isomorphism property holds,  $\langle M, s \rangle \models \varphi$  iff  $\langle M', s' \rangle \models \varphi$ . Therefore, by the definitions of  $s$  and  $s'$ ,  $M \models C[b_1, \dots, b_n]$  iff

---

<sup>24</sup>More precisely: the nonlogical terms in  $\varphi$  include *at most* variables. Variables can be considered as either nonlogical terms, or as neither logical nor nonlogical - which is how Sher treats them, see (Sher, 1991, p. 84). That decision does not affect our proof.

### CHAPTER 3. MODELS AND LOGICAL CONSEQUENCE

$M' \models C[b'_1, \dots, b'_n]$ , thus  $\langle b_1, \dots, b_n \rangle \in f_C(M)$  iff  $\langle b'_1, \dots, b'_n \rangle \in f_C(M')$ , as required.

The right to left direction is proved by induction, the details of which we leave to the reader.<sup>25</sup>

---

<sup>25</sup>For more details of the proof of the equivalence result, see (Sagi, 2011).

## Chapter 4

# Criteria for Logical Terms: Varieties of Invariance

### 4.1 Invariance Criteria for Logical Terms: The Basics

Much of the recent and contemporary work on logicality concerns a class of criteria for logical terms which I have labeled “invariance criteria”. The idea is that due to their formal nature, logical terms are indifferent to certain properties of models. Each invariance criterion sets the properties of models to which logical terms, according to that criterion, may be sensitive. Any two models that are the same with respect to those properties, should give the same interpretation to the logical terms. There is wide agreement that logical terms should be indifferent to differences between objects. This idea yields *Tarski’s thesis* and the *Tarski-Sher thesis*, formulated below. Some writers think these criteria are too weak, because they allow logical terms that are sensitive to the cardinality of the domain (see Section 4.3), and have thus proposed more restrictive alternatives.

The criterion for logical terms as invariant under isomorphisms was presented in Chapter 3 in the discussion of the work of Gila Sher and Stewart Shapiro. This chapter consists

of a survey of the work of three additional writers: Vann McGee, Solomon Feferman and Denis Bonnay. Each argues for a criterion of invariance under some kinds of transformations. To facilitate the reading of this chapter, the basic notions of permutation invariance and isomorphism invariance are defined and explained.

### 4.1.1 Permutation invariance

In 1966 Tarski delivered a lecture in which he proposed a criterion for logical notions ((Tarski, 1986), see also (Tarski & Givant, 1987)). His criterion of permutation invariance is presented as a continuation of Klein’s *Erlanger Programm*. Klein distinguished between the notions of various disciplines in geometry according to the kind of transformations under which they are invariant. For instance, size and location do not matter for the objects of Euclidean geometry, and only the ratio between distances make a difference. Thus, the notions of Euclidean geometry are invariant under similarity transformations (transformations which preserve the ratio between distances). Analogously, the notions of affine geometry are invariant under affine transformations, and the notions of topology are invariant under continuous transformation (and more specifically, under homeomorphisms).

The idea is then extended by Tarski to the realm of logic. The idea is that logic is the most general discipline, and therefore its notions should be invariant under any transformation whatsoever. Now note that notions, as opposed to terms, are mathematical entities in some set-theoretic hierarchy. Tarski’s criterion can then be modified to apply to terms, as those that denote logical notions (Tarski & Givant, 1987, p. 57). Let  $U$  be some non-empty set, the ‘universe of discourse’. From  $U$  we can construct derivative universes of higher types. The union of such universes gives us the set-theoretic hierarchy with the members of  $U$  as ur-elements. Let  $\pi$  be a permutation on  $U$  (a *transformation* in Tarski’s terminology). Then for each derivative universe  $\tilde{U}$  there is a permutation  $\tilde{\pi}$  naturally derived from  $\pi$ , such that for  $a \in U \cap \tilde{U}$ ,  $\tilde{\pi}(a) = \pi(a)$  and for each set  $A \in \tilde{U}$ ,  $\tilde{\pi}(A) = \{\tilde{\pi}(B) : B \in A\}$ .

## CHAPTER 4. VARIETIES OF INVARIANCE

A notion  $o$  belonging to  $\tilde{U}$  is invariant under permutations if for every permutation  $\pi$  on  $U$ ,  $\tilde{\pi}(o) = o$ .

Now we are able to state the criterion of permutation invariance, which will also be referred to henceforth as *Tarski's thesis*:

### **Tarski's thesis (logical notions)**

Given a basic universe  $U$ , a member  $o$  of any derivative universe  $\tilde{U}$  is logical if it is invariant under permutations.

### **Tarski's thesis (logical terms)**

Given a basic universe  $U$  and a language  $L$ , a term  $a$  in  $L$  is logical if it denotes a member of a derivative universe  $\tilde{U}$  that is invariant under permutations.

(See (Tarski & Givant, 1987, p. 57))

Note that identity and the quantifiers, standard (as  $\forall$  and  $\exists$ ) and generalized (as  $\exists_\infty$ : there are infinitely many) pass the test of Tarski's thesis. Truth-functional connectives need to be somehow accommodated in order to be considered permutation invariant. This can be done in various ways, see (Sher, 2008).

### **4.1.2 Isomorphism invariance**

Tarski's thesis defines logicality relative to a given domain. Contemporary model theory, however, employs multiple domains. One can extend Tarski's thesis to deal with multiple domains by requiring that a logical notion be permutation invariant in all domains of models, and that logical terms denote such notions. But then there are other maps between domains that are not permutations and might be considered, namely those between different domains.

Gila Sher has suggested that since logical terms are formal, they denote only structural objects and properties, isomorphisms are the appropriate functions to use. Sher's criterion for logical terms can be formulated in terms of isomorphisms or of bijections between



domains. Here I shall formulate the criterion using bijections.<sup>1</sup>

Notions in this context are functions from models to the set-theoretic hierarchy with the models' domains as ur-elements. Let  $o$  be a notion and  $a$  be a term in a language  $L$ . We shall say that  $a$  denotes  $o$  if  $o(M) = I(a)$  for every model  $M = \langle D, I \rangle$  for  $L$  with domain  $D$  and interpretation function  $I$ . Now let  $M = \langle D, I \rangle$  and  $M' = \langle D', I' \rangle$  be models, and let  $f$  be a bijection from  $D$  onto  $D'$ . As with permutations before,  $f$  can be naturally extended to apply any derivative universes  $\tilde{D}$  and  $\tilde{D}'$  of  $D$  and  $D'$  if they are of the same type.  $o$  is invariant under bijections if for any models  $M = \langle D, I \rangle$  and  $M' = \langle D', I' \rangle$  and for any bijection  $f$  between  $D$  and  $D'$  extended in the natural way to all derivative universes of  $D$  and  $D'$ ,  $f(o(M)) = o(M')$ . Now we are ready to formulate the Tarski-Sher thesis

### **The Tarski-Sher thesis (logical notions)**

A notion  $o$  is logical if it is invariant under bijections.

### **The Tarski-Sher thesis (logical terms)**

A term  $a$  in a language  $L$  is logical if it denotes a logical notion.

According to the Tarski-Sher thesis identity as well as standard and generalized quantifiers are logical. The main difference between the Tarski-Sher thesis and Tarski's thesis is that some terms that are logical by the latter are not logical by the former. Such is "wombat quantification",  $Q$ , inspired by McGee (McGee, 1996, p. 575): on domains containing wombats,  $Q$  is interpreted as the universal quantifier, and on domains without wombats  $Q$  is interpreted as the existential quantifier.  $Q$  is invariant under permutations, since on

---

<sup>1</sup>Bijections are a subclass of isomorphisms: they are isomorphisms between sets with no consideration of internal structure. Once more structure is added, the notion of isomorphism needs to be employed. Note also that *isomorphism* is also used to describe a relation between models in addition to a function (two models are isomorphic iff there is an isomorphism from one of them to the other). Sher uses isomorphisms in her formulation in (Sher, 1991).

## CHAPTER 4. VARIETIES OF INVARIANCE

any given domain it is interpreted as just one of the quantifiers, regardless of how the domain might be permuted.  $Q$  is not invariant under bijections: there may be isomorphic domains, one containing wombats and one lacking them, and on these two domains  $Q$  will be interpreted differently.

The Tarski-Sher thesis is widely accepted in the literature (e.g. (McCarthy, 1981; van Benthem, 1989; Sher, 1991, 1996; McGee, 1996)). Sher imposes some further conditions on logical terms, her full criterion is given in her book (Sher, 1991, pp. 54f).

## 4.2 McGee's Justification of Permutation Invariance

### 4.2.1 Summary

Vann McGee (McGee, 1996) proposes an interesting and novel argument for Tarski's thesis (by which a notion is logical if and only if it is invariant under all permutations of the domain) and the more general Tarski-Sher thesis (by which logical notions are invariant under isomorphisms (see Section 4.1)).<sup>2</sup>

We start with Tarski's thesis, which is easier to work with. McGee supports Tarski's thesis as it is applied to operations (which are akin to Tarski's *notions*), while as regards to *terms* (*connectives* in his terminology) permutation invariance of the extensions of terms is necessary but not sufficient for logicality.

McGee accepts Tarski's reasoning for the necessity of the condition of permutation invariance for logicality. He adds to it a novel argument for the sufficiency of permutation invariance for the logicality of operations (given a fixed domain) that relies on the characteristics of the infinitary language  $L_{\infty\infty}$ . McGee proves that an operation is invariant under permutations of the domain if and only if there is some formula in  $L_{\infty\infty}$  that describes it. Since, as McGee claims, intuitively,  $L_{\infty\infty}$  is purely logical (its building blocks are all logical and therefore anything describable by it will be logical), all permutation invariant operations should be considered as logical relative to the given domain. McGee then discusses the possibility of extending the argument to deal with varying domains, as well as proper-class domains. As regards logical *connectives*, McGee uses counterexamples to argue that permutation invariance is insufficient, and proposes a way to strengthen the criterion. In this section I present McGee's main argument and discuss the presuppositions that are involved.

---

<sup>2</sup>In this chapter we use *invariance under isomorphisms* (the terminology used by Sher) and *invariance under bijections* (the terminology used by McGee) interchangeably, since they amount to the same thing in the contexts we use them.

### 4.2.2 Connectives and Operations

McGee divides the discussion of logicality between operations and connectives. Connectives are syntactic items in a language, and operations are mathematical entities that may be described by connectives. It appears that for McGee the issue of the logicality of connectives is primary, and the logicality of operations is dealt with as a step in the way. The distinction between logical and nonlogical connectives is particularly important, according to McGee, since it plays a major role in Tarski's theory of truth and logical consequence (Tarski, 1936), (McGee, 1996, p. 569).

A criterion for the logicality of operations will not immediately provide a criterion for the logicality of connectives, since one and the same operation might be described by both a logical and a nonlogical connective. McGee gives *unicorn negation* as an example for a nonlogical connective that is co-extensive with a logical one:

$$\mathcal{U}\phi =_{def} (\neg\phi \wedge \text{there are no unicorns})$$

(p. 569)

$\mathcal{U}$  is co-extensive with plain negation, that is, they both describe the same operation. But obviously, according to McGee, unicorn negation should not be considered as a logical connective. McGee thus sets the relation between the logicality of operations and connectives in the following way: any operation that is described by a logical connective is logical (p. 570). For a connective to be logical, it has to follow from its meaning that the operation it describes satisfies the criterion for logicality (see p. 578). We will return to this last point later on.

McGee has a useful way of treating all connectives—all the items in a language that are candidates for logicality—in a uniform way. Each connective describes an operator of some arity on sets of assignments. Let  $S$  be a nonempty set, serving as the domain, and let  $V$  be a set of variables. For any cardinal  $\iota$ , a  $\iota$ -ary operator takes as input a sequence of  $\iota$  sets of  $V$ -assignments for  $S$  and returns a set of  $V$ -assignments for  $S$ . Truth-functions,

predicates and quantifiers can all be described as operating in this way. In this setting, the connectives negation, disjunction and conjunction are viewed as describing operations of complementation, union and intersection of sets of assignments. Existential quantification is viewed as describing cylindrification.

Formulas can also be considered as connectives, and actually, McGee’s argument refers only to formulas. It is natural to view a formula as representing simply a set of  $V$ -assignments, but McGee uses formulas to describe all possible *operations*. We will see shortly how this is done.

### 4.2.3 Tarski’s thesis

Tarski’s thesis (Tarski, 1986), in McGee’s terminology, is the following:

The logical operations on a domain are those operations that are invariant under all permutations. (McGee, 1996, p. 567)

McGee agrees with the Tarskian reasoning supporting one direction of Tarski’s thesis. Permutation invariance, according to McGee, is obviously a *necessary* condition for logicality:

Any operation which is disturbed by a permutation must somehow discriminate among individuals in the domain, and any consideration which discriminates among individuals lies beyond the reach of logic, whose concerns are entirely general. (*ibid*)<sup>3</sup>

---

<sup>3</sup>It would be more accurate to attribute this type of reasoning to Tarski with conjunction of Sher. An interesting fact on Tarski work on the subject is usually ignored: The contention that logic should not be affected by particular features of objects appears in (Tarski, 1936), where Tarski discusses logical consequence, and not in (Tarski, 1986), where he discusses logical notions. The insensitivity to particular objects is brought in the first article as a feature of the formality of logic, without claiming that *logical terms* are responsible for this feature, and without hinting that this might serve as a condition for logical terms. In the latter article, Tarski proposes his criterion of invariance without this kind of philosophical motivation.

## CHAPTER 4. VARIETIES OF INVARIANCE

McGee acknowledges that Tarski has some justification for the sufficiency of permutation invariance, and that the consideration that situating logic in Klein’s *Erlangen Program* makes permutation invariance a very natural choice. McGee, nonetheless, proposes an argument that gives further support to Tarski’s thesis. He then shows how this argument may extend to deal with logical terms independently of a specific domain, replacing permutations with bijections, to support what we earlier termed the *Tarski-Sher thesis*.

### 4.2.4 The argument from $L_{\infty\infty}$

McGee’s argument for the sufficiency of permutation invariance, in very general lines, is the following. Let  $O$  be some permutation invariant operation. Then it can be described by a connective (specifically, a formula) which is unmistakably logical, as it is constructed from obviously logical primitive connectives. Since any operation described by a logical connective is logical, the logicity of the formula implies the logicity of the operation described.

In order to evaluate McGee’s argument, we need to look into the language  $L_{\infty\infty}$  and the theorem McGee proves. We will take special notice of McGee’s claim that any connective defined in  $L_{\infty\infty}$  is intuitively logical.

#### The language $L_{\infty\infty}$

As before, let  $S$  be a non-empty set. For each cardinal  $\iota$ , we construct a version of the language  $L_{\infty\infty}$  which will provide an appropriate formula for each  $\iota$ -ry operation on

---

He acknowledges that the criterion yields notions of “a very general character” (Tarski, 1986, p. 149), and moves on to discuss the consequences of his definition. It seems that his main justification for his criterion is that it agrees in general with logical practice (*ibid* p. 150), and that it situates logic as a mathematical discipline in a natural way, extending Klein’s *Erlangen program*.

In Sher’s work we can finally find the connection between logical terms and indifference to particular features of objects, albeit inspired by Tarski (Sher, 1991, 1996).

assignments on the domain  $S$ .

- The *predicates* of the language will be  $\{P_i : i < \iota\}$ .
- As noted before,  $V$  is a set of variables. In addition, we have a set of variables  $W$  that includes  $V$  and also at least  $\text{card}(S) + 1$  variables not in  $V$  that will be used for quantification.
- A *substitution* is a function from  $V$  to  $W$ .
- The *atomic formulas* are expressions of the form  $P_i s$ , where  $P_i$  is a predicate and  $s$  a substitution, or of the form  $v = w$  where  $v$  and  $w$  are variables.

Complex formulas:

- if  $\phi$  is a formula,  $\neg\phi$  is a formula.
- If  $\Phi$  is a set of formulas, its disjunction  $\bigvee \Phi$  is a formula.
- If  $\phi$  is a formula and  $U$  a set of variables,  $(\exists U)\phi$  is a formula.
- Conjunction and universal quantification are non-primitive, and are defined in the usual way.

The model theory for this language is the standard one for first order languages with some modifications. For instance, predicates are of no specific arity here. A predicate's extension in a model is simply a set of assignments.

The recursive conditions for satisfaction are straightforward given the definition of formulas in  $L_{\infty}$ —see (McGee, 1996, p. 571)).

Now let  $\langle A_i : i < \iota \rangle$  be a sequence of sets of assignments. We define  $\mathcal{M}_{\langle A_i : i < \iota \rangle}$  to be the model that assigns  $A_i$  to  $P_i$  as its extension, for each  $i < \iota$ .

The *operation described by a formula  $\phi$*  is the  $\iota$ -ry operation that takes  $\langle A_i : i < \iota \rangle$  to  $\{\sigma : \sigma \text{ satisfies } \phi \text{ under } \mathcal{M}_{\langle A_i : i < \iota \rangle}\}$ .

## CHAPTER 4. VARIETIES OF INVARIANCE

An operation  $O$  is *invariant under permutations* if for any permutation  $\pi$  on  $S$ , extended naturally to assignments, sets of assignments, and sequences of sets of assignments on  $S$ , and for any sequence  $\langle A_i : i < \iota \rangle$ ,  $\pi(O(\langle A_i : i < \iota \rangle)) = O(\pi(\langle A_i : i < \iota \rangle))$ .

**Theorem 1** *An  $\iota$ -ry operation is invariant under permutations if and only if it is described by some formula of  $L_{\infty\infty}$  (McGee, 1996, p. 572).*

I will not reproduce the full proof here. The right to left direction is proved by induction on formulas. The left to right direction is proved by construction a formula given a permutation invariant operation. It is the latter that serves McGee’s main point, that any permutation invariant operation is described by an obviously logical connective.

The presentation in this section, up to this point, followed McGee almost word to word. Our job now will be to assess the strength of McGee’s argument by exposing some of the assumptions in using Theorem 1.

For the sake of his argument, McGee needs *all* formulas of  $L_{\infty\infty}$  to describe an intuitively logical operations. That is since *all* formulas of  $L_{\infty\infty}$ , by the theorem, describe a permutation invariant operation. Note that this includes *atomic* formulas as well. It might seem that, at least at first glance, we would not conceive of an atomic formula such as  $P_i s$  as describing a logical operation. It is nowhere near a logical truth—what does it mean that the operation that it describes is logical? To resolve this quandary, we refer back to the definition of atomic formulas, and to the definition of the operation described by a formula. First, we note that  $s$  here is a substitution, i.e. a function from  $V$  to  $W$ : a mere change of variables. Secondly, we note that  $\phi$  describes an operation  $O$  *given an assignment to the predicates*. That is, the operation described by  $P_i$  is not that of predicating  $P_i$  of some individuals, but it is the operation of *substitution*, set by  $s$ , on a given set of assignments (those that were assigned to  $P_i$ ).

Note also, that even though McGee talks of disjunction, conjunction and negation as



describing union, intersection and complementation, the operations dealt with here do not include those set-theoretic operations *per se*, since none of the operations McGee describes are unary or binary. Disjunctions, conjunctions and negation have not been assigned operators in McGee's definitions of  $L_{\infty\infty}$ —only formulas have been assigned operators. And given a formula  $\phi$ , the operation described by  $\neg\phi$  will not simply produce the complement of the set of assignments satisfying  $\phi$  from those satisfying  $\phi$ . The input, remember, is the sequence  $\langle A_i : i < \iota \rangle$  of sets of variable-assignments: an assignment to all predicates of the language. This sequence nonetheless determines the set of assignments satisfying  $\phi$ , as it is constructed from these predicates. But the operation described by  $\neg\phi$  is not literally that of complementation. I take it that this last point makes little difference to the strength of McGee's argument, but it is crucial in understanding the theorem he uses in the argument.

Now McGee's argument relies on the following claim:

(I) Any connective constructed in  $L_{\infty\infty}$  is logical.

This claim follows from the following two assumptions:

(II) All the primitive connectives in  $L_{\infty\infty}$  are logical.

(III) Any connective constructed from logical connectives is logical.

I'd now like to examine (II) in some detail. I will leave (III) aside, assuming it is reasonable enough.

McGee justifies (II) by enumerating the primitive connectives in  $L_{\infty\infty}$  and appealing to intuition. The primitive connectives of  $L_{\infty\infty}$  are:

- the identity relation
- substitution of variables
- finite or infinite disjunction

## CHAPTER 4. VARIETIES OF INVARIANCE

- negation
- existential quantification, perhaps with regard to an infinite block of variables (p. 568).<sup>4</sup>

It should be stressed that the connectives in the list above quoted from McGee are not assigned operations in his definitions, since he deals only with operations described by *formulas*. McGee uses the intuitive logicity of the connectives above to argue for the logicity of the formulas they make up, and uses formulas to make his point on the logicity of all permutation invariant operations.

Now it seems to me that the intuitive logicity of this or that connective is not much an issue for debate, but rather more of a subjective matter that you either share or don't share with the author. I interpret McGee's appeal to intuitions as saying that he is taking it for granted that the above connectives are logical, that he assumes that his readers will have no trouble doing the same, and that there is no need to present any further argument for the matter. This is fair enough, as long as it is accepted that some readers might not share these intuitions. I admit that I am somewhat at loss when it comes to intuitions of logicity of this or that connective. I will not take a stand here on whether the connectives McGee lists are (intuitively) logical (this is a question I dismiss in Part II). However, I would like to highlight some features of these connectives that might impugn the intuitions that might be involved.

The infinitary nature of  $L_{\infty\infty}$  involves more set theory than appears at first, that sneaks in with the infinite disjunctions and the infinite quantification. For instance, disjunction is performed by specifying a set of formulas. The set specified can be *any set*, specified in any

---

<sup>4</sup>To be accurate, when enumerating the above items, McGee refers to them as intuitively logical *operations*. But this is in the beginning of the article, before he distinguished between operations and connectives. Clearly, those items are *connectives* on the terminology introduced later on: they are syntactic items where the associated operations include union, complementation etc.

way.<sup>5</sup> If our set theory admits that, we can specify the set  $\mathcal{A}$  of all formulas  $P_i$  for all even finite  $i$  such that  $i$  satisfies Goldbach's Conjecture, that is, for  $i$ 's that are a sum of two primes. Or the set  $\mathcal{B}$  of all formulas written in North America in 1996, if such a set exists. Disjunction over any sets allowed in our set theory is allowed in  $L_{\infty\infty}$ . Hence, claiming that the connectives of  $L_{\infty\infty}$  are logical includes the claim that the formulas  $\bigvee_{i \in \mathcal{A}} P_i$  and  $\bigvee \mathcal{B}$  are logical connectives—although they make reference to sets that are specified through substantial mathematics and empirical facts. The same goes for conjunction, and similarly for existential and universal quantification. One might view this as unproblematic, but still, it is important to note that the *intuitive logicity* of such connectives is appealed too, and one should run their intuitions on such cases as those mentioned as well. I predict that while many readers will find the connectives in the list presented above intuitively logical at first, their attitude will change when realizing what the infinite disjunctions and infinite quantification actually involve.

McGee could run his argument, using Theorem 1, without letting *all* disjunctions into  $L_{\infty\infty}$ . He might restrict himself to a modest set theory where  $\mathcal{A}$  and  $\mathcal{B}$  are not allowed. Let us see what the proof actually requires. In his proof for Theorem 1, for the left to right direction, McGee takes an arbitrary operation  $O$ , and shows that if it is permutation invariant then there is a formula describing it in  $L_{\infty\infty}$ . McGee uses there a conjunction ranging over all predicate assignments  $\langle A_i : i < \iota \rangle$  and a disjunction over all assignments belonging to  $O(\langle A_i : i < \iota \rangle)$  (p. 572). Further, McGee enumerates the domain  $S = \langle s_\alpha : \alpha < \kappa \rangle$  assuming  $\text{card}(S) = \kappa$ , specifies a variable  $x_\alpha$  for each  $s_\alpha$ , and uses universal quantification, disjunction and conjunction that all rely on this enumeration (p. 572f). These connectives should all be intuitively logical in McGee's account. To repeat, I do not wish to argue here *against* the logicity of such connectives, but merely expose what McGee's argument involves. I am leaving it to the reader to decide whether the connectives

---

<sup>5</sup>Note that since McGee concerns himself with the intuitive logicity of *connectives* here—so not only their extensions but also their meaning and the way they are specified matter to their logicity.

## CHAPTER 4. VARIETIES OF INVARIANCE

used in McGee's proof are all intuitively logical: whether their logicality is something that can reasonably be taken for granted. In case they are not, McGee's argument is unsound, and does not provide any further support to Tarski's thesis.

Further criticism in this line is put forward by Feferman (1999). Feferman claims that  $L_{\infty\infty}$  has the means to express second-order quantification, and can therefore express substantial mathematical claims such as the continuum hypothesis (CH) deeming them logically determinate (Feferman, 1999, p. 37f).<sup>6</sup> Feferman criticises the resulting assimilation of logic to set theory. Now it is not enough that a proposition be expressible in  $L_{\infty\infty}$  for it to be logically determinate in this language. Obviously, not *all* sentences in  $L_{\infty\infty}$  are logical truths or logical falsehoods:  $\forall x(P_0x)$  is an example. One should thus distinguish between *the operation described* by a formula in  $L_{\infty\infty}$ , which is always logical under the Tarski-Sher thesis, and *the proposition expressed* by a formula in  $L_{\infty\infty}$ , which should not always be logically determinate. However, Feferman adds that second-order quantification turns out to be permutation invariant, that is, a logical operation by Tarski's thesis.

It might seem that Feferman invites us to add the claim that CH can be expressed by purely logical vocabulary in  $L_{\infty\infty}$  and therefore turns out to be logically determinate. But this reasoning is invalid, at least when we allow the domain to vary: *there are at least two objects*,  $\exists x\exists y\neg(x = y)$  is a sentence in  $L_{\infty\infty}$  constructed by purely logical vocabulary but is not logically determinate.

So what exactly is the claim made by Feferman on statements such as CH? Even though we cannot say that any sentence constructed from purely logical vocabulary has the same truth value in all domains, we can still say that it has the same truth value in all domains of the *same cardinality*. This is a direct result of the *isomorphism property* discussed in Chapter 3. As we have seen in that chapter, the isomorphism property is equivalent to Sher's criterion (i.e. the Tarski-Sher thesis), so if all logical terms in a language satisfy Sher's criterion (as they do in  $L_{\infty\infty}$ ), the isomorphism property holds, and the truth in a

---

<sup>6</sup>For a complete formalization of CH in second-order logic, see (Shapiro, 1991, pp. 100-105).

model of sentences constructed by purely logical vocabulary becomes a matter of the size of the domain. Indeed, we can express CH relative to domains of size  $\kappa$ , for each cardinal  $\kappa$ . Each such relativized  $\text{CH}_\kappa$  will be true only in models with a domain of size  $\kappa$ , and will be given the same truth value in all such models. This should be enough to substantiate Feferman's claim that the Tarski-Sher thesis assimilates logic to set theory.

Leaving aside Feferman's method for accommodating second-order quantification in  $L_{\infty\infty}$ , we may note that the ability to quantify over arbitrary sets of variables leads the way to second-order quantification (see also (Barwise, 1969, p. 227)). But in fact, the formulation of the continuum hypothesis does not require the full machinery of second-order quantification. Let us see how CH can be formalized directly. The statement we shall attempt to formalize is:

CH: There is no set  $S$  such that  $\aleph_0 < \text{Card}(S) < 2^{\aleph_0}$ .

Now let  $c$  be the first ordinal the size of the continuum:  $c = \min\{\alpha : \alpha \text{ is an ordinal, } \text{card}(\alpha) = 2^{\aleph_0}\}$ . The formalization of CH in  $L_{\infty\infty}$  will be:

$$\text{CH}_{L_{\infty\infty}}: \forall(\{x_\alpha : \alpha \in c\}) \bigwedge_{S \subseteq c} \left[ \left( \forall(\{y_\alpha : \alpha \in c\}) \bigwedge_{\alpha \neq \beta \in c} y_\alpha \neq y_\beta \rightarrow \bigvee_{\alpha \in c} \bigwedge_{\beta \in S} y_\alpha \neq x_\beta \right) \rightarrow \left( \exists(\{z_\alpha : \alpha \in \omega\}) \bigwedge_{\alpha \in S} \bigvee_{\beta \in \omega} x_\alpha = z_\beta \right) \right]$$

In words: For each set that is embeddable in  $c$ , if  $c$  is not embeddable in it, then  $\omega$  is embeddable in it. In terms of cardinality: For each set of cardinality less or equal to  $2^{\aleph_0}$ , if its cardinality is strictly less than  $2^{\aleph_0}$ , then its cardinality is less or equal to  $\aleph_0$ . Thus,  $\text{CH}_{L_{\infty\infty}}$  expresses CH. If the basic connectives of  $L_{\infty\infty}$  seem intuitively logical to some at first, I suspect that seeing how they can be utilized in formalizing substantial set-theoretic claims might alter this intuition.

### 4.2.5 Concluding remarks

In the article this section is devoted to, McGee presents a novel, impressive argument in support of Tarski's thesis that the logical operations are those that are permutation invariant. I have summarized the argument, going through most of the technicalities involved. I discussed McGee's assumption (II), that the primitive connectives of  $L_{\infty\infty}$  are all intuitively logical. I pointed out that in this language, *any set* can be specified when disjunction and other connectives are employed, and left it for the reader to decide whether this has any effect on their intuitions.

Some further issues raised in McGee's article are worth mentioning. McGee extends his argument to deal with logical operators over varying domains. The extended criterion for logicality of a term in varying domains McGee endorses is that of the Tarski-Sher thesis, and the appropriate theorem is: an operation  $O$  across domains is invariant under bijections iff, for each nonzero cardinal  $\kappa$ , there is a formula  $\phi_\kappa$  of  $L_{\infty\infty}$  that describes the action of  $O$  on domains of cardinality  $\kappa$ . Now for each cardinal  $\kappa$  take the appropriate formula  $\phi_\kappa$ . The conjunction of the  $\phi_\kappa$ s, if one allows a proper class of conjuncts, will be a formula that completely describes the action of  $O$  (McGee, 1996, p. 576). The remarks made earlier on the intuitive logicality of the basic connectives McGee uses are still valid here, and the main point is accentuated owing to the use of a proper-class-conjunction. McGee further deals with operations of higher types and non-set domains, which we leave out of this survey.

Another issue is that of a criterion for logical *connective*. Recall that permutation invariance is for McGee a criterion for logical *operations* (even though he uses the logicality of some connectives on the way). But not every connective that describes a logical operator is itself logical : McGee gives *unicorn negation* as an example (see Section 4.2.2 and (McGee, 1996, p. 569)). McGee claims that *necessary* permutation invariance is still not a strong enough condition (by which a connective is logical if it is necessarily permutation invariant), giving the example of *H<sub>2</sub>O-negation*:

$$\mathcal{H}\phi =_{def} (\neg\phi \wedge \text{water is } H_2O)$$

If  $\mathcal{H}$  were a logical connective, claims McGee, “ $(\neg 0 = 0 \vee \mathcal{H}\neg 0 = 0)$ ” would be a logical truth, although it entails “water is  $H_2O$ ”, which clearly isn’t (p. 578). Thus, McGee proposes the following conjecture: “A connective is a logical connective if and only if it follows from the meaning of the connective that it is invariant under arbitrary bijections” (*ibid*).

In Chapter 5 I discuss arguments against Sher’s criterion using such counterexamples, and claim that these arguments rely on a confusion between metalanguage and object-language. I concentrate mainly on arguments given by McCarthy and Hanson (see specifically f.n. 13 mentioning McGee). To apply here the criticism in short, let us observe the definition of  $\mathcal{H}$ . One option is that this definition is given in the *metalanguage*. In that case, I agree with McGee that  $\mathcal{H}$  describes the same operation as standard negation. However, I disagree that “ $(\neg 0 = 0 \vee \mathcal{H}\neg 0 = 0)$ ” entails “water is  $H_2O$ ”—at least not when *entails* is understood as *a logical consequence of* in any standard sense. The first sentence quoted is in the object-language, and the second one is in the metalanguage. It would only make sense to speak of a relation of entailment if both sentences are considered in the same language. We should reasonably assume that that would be the metalanguage. But it is not clear how the entailment relations between sentences in the metalanguage should affect their logical truth as sentences in the object language (see Chapter 5.4), and McGee does not help us out. If, on the other hand, the definition is understood as stated in the object-language, I see no reason why  $\mathcal{H}$  defines the same operator as negation. Obviously, McGee would not consider *water* and  $H_2O$  to be logical connectives. Thus, their interpretations would vary from model to model, and in some models would not be identical. In those models,  $\mathcal{H}\phi$  would be false, regardless of  $\phi$  (see Chapter 5.4 p. 101 for further elaboration of this point).

### 4.3 Feferman: Invariance under Homomorphisms

#### 4.3.1 Summary

In this section we look into Feferman’s “Logic, Logics and Logicism” (Feferman, 1999). Feferman rejects Tarski’s thesis as well as the more general Tarski-Sher thesis. Feferman diagnoses the problems he detects in those theses as resulting from the use of permutations and bijections/isomorphisms. Feferman proposes an alternative thesis with the same structure: logical operations are still those that are invariant under some sort of mapping—where homomorphisms fulfill the role as the type of mapping to look at.

The main benefits of the new thesis, according to Feferman, are that it sets a better boundary between logic and set theory, and that it captures the idea of a term denoting the same operation over different models. Feferman further supports his criterion by proving that an operation is definable from the first-order predicate calculus without identity just in case it is definable from homomorphism invariant monadic operations. A perhaps unwanted outcome of Feferman’s criterion is that identity doesn’t satisfy it, and needs to be added stipulatively if it is to be considered as a logical term.

In the latter part of the section I discuss the implications of Feferman’s choice of formal framework. I indicate that the criterion of homomorphism invariance yields the desired results in Feferman’s framework and not in others. The use of Feferman’s criterion and the philosophical implications drawn from it should be accordingly restricted.

#### 4.3.2 Against the Tarski-Sher thesis

In the previous section we have seen, due to McGee, that the operations invariant under permutations given some domain are those definable in  $L_{\infty\infty}$ . Looking at varying domains, McGee points out that operations invariant under bijections between any two domains are definable by a proper class conjunction of formulas in  $L_{\infty\infty}$ .

Feferman has three main concerns with the Tarski-Sher thesis. The first is that it as-



simulates logic to mathematics, or, more accurately, to set theory. This concern was raised in the previous section discussing McGee’s argument. There I pointed out that the basic connectives McGee refers to as “intuitively logical” involve substantial set theory. Feferman’s criticism, as we have seen, is directed to the availability of second-order quantification in  $L_{\infty\infty}$ . Feferman claims that  $L_{\infty\infty}$  has the means to express second-order quantification, and can therefore express mathematical claims such as the continuum hypothesis, deeming them logically determinate (Feferman, 1999, p. 37f).

Secondly and relatedly, Feferman points out that many terms that are logical on the Tarski-Sher thesis are not set-theoretically robust. By this he means that their extension varies between different interpretations of set theory, so it is not *absolute*. For instance, the quantifier  $\exists^{>\aleph_0}$  (“there are uncountably many”) denotes a non-absolute property (this a direct result of the Löwenheim-Skolem theorem). Feferman suggests imposing absoluteness as a necessary condition on logical terms if set theory is taken as the background theory for logic, to keep logic independent of the actual extent of the set-theoretic universe ((Feferman, 1999, p. 38), cf. (Bonnay, 2008) discussed in the next section).

Thirdly, Feferman criticizes the Tarski-Sher thesis for allowing terms that “behave differently” in domains of different sizes. The Tarski-Sher thesis only requires that logical terms be invariant under *isomorphic* structures – which means that they can vary considerably between structures of different cardinalities. This means that they may be sensitive to the cardinality of the domain. As a result, the Tarski-Sher thesis allows “domain-sensitive quantifiers”. An instance of such a quantifier is the second-level predicate  $C$  that “behaves” like the universal quantifier in models with domains of certain cardinalities (e.g., odd finite ones) and as the existential quantifier in all the rest. Such a quantifier is very unnatural, as it seems to change its meaning completely between models of different cardinalities (see also (Hanson, 1997)). Feferman contends that this is a major disadvantage of the thesis, and attempts to overcome the problem with his proposed criterion.

### 4.3.3 Feferman's criterion

To overcome the problems with the Tarski-Sher thesis, Feferman proposes a modified criterion for logical terms. The antecedent of the Tarski-Sher thesis refers to *isomorphic models* (meaning simply that there is a bijection between their domains, i.e. they are equinumerous). The main objections to the Tarski-Sher thesis are that it is too permissive, it rules too many operations as logical. If one is to keep the basic idea of the Tarski-Sher thesis, the key to addressing these objections is by changing the relation of isomorphism to a more *inclusive* relation, so that more pairs of models would be required to satisfy the condition of invariance. Feferman proposes the relation of homomorphism. A *homomorphism* (in Feferman's usage) is simply a surjective map. While every isomorphism is a homomorphism, the reverse is not true: a homomorphism need not be injective. This means that models of different cardinalities can be compared.

Before presenting Feferman's criterion, we should point out that Feferman uses a functional finite type framework. This means that he defines all operations as functions. The basic types are 0 (the type of the domain of individuals) and  $b$  (the type of truth values). The functional type symbols are generated from 0 and  $b$  by formation of  $\tau = (\tau_1, \dots, \tau_n \rightarrow \sigma)$  where  $\tau_1, \dots, \tau_n, \sigma$  are functional type symbols. For each model  $M$ , we associate with each type  $\tau$  a universe  $M_\tau$  over  $M$ .  $M_0$  is the domain of  $M$  and  $M_b = \{T, F\}$ . For each type  $\tau = (\tau_1, \dots, \tau_n \rightarrow \sigma)$ ,  $M_\tau$  consists of functions from  $M_{\tau_1} \times \dots \times M_{\tau_n}$  to  $M_\sigma$ .

Let  $h : M_0 \rightarrow M'_0$  be a homomorphism from the domain of  $M$  onto the domain of  $M'$ .  $h$  can be extended to higher types, so as to apply to all operations. An operation  $O$  is *homomorphism invariant* if for any two models  $M$  and  $M'$  and a homomorphism  $h$  between them and extended to higher types,  $h(O) = O$ . According to Feferman, "there is a natural sense in which *operations  $O$  invariant under homomorphisms are logical form preserving, if one ignores equality*, at least for propositional operations  $O$  of type level 2 with proposition function arguments" (Feferman, 1999, p. 40). Indeed, most of the logical terms

of standard logic are homomorphism invariant. Identity of individuals is an exception: since a homomorphism can take two individuals in one domain to the same individual in another, identity is not preserved (see Section 4.3.4). We shall refer to homomorphism invariance as *Feferman's criterion* (even though he does not endorse this criterion wholeheartedly, and suggests possible modifications).

Before moving on to Feferman's definability result, let us look at some examples to get an idea of the logic of homomorphism invariant operations.

The following operations are homomorphism invariant:

- Negation, as a type  $(b \rightarrow b)$  operation.
- Conjunction, as a type  $(b^2 \rightarrow b)$  operation, as well as the rest of the truth-functional operations.
- The existential quantifier,  $\exists$ , as a  $((0 \rightarrow b) \rightarrow b)$  operation.
- The *wellfoundedness quantifier*,  $W$ , as a  $((0^2 \rightarrow b) \rightarrow b)$  operation, applied to binary relations between individuals, defined by:

$$W^M(r) = \begin{cases} T & \text{if } (\forall f : \omega \rightarrow M_0)(\exists n \in \omega)r(f(n+1), f(n)) = F \\ F & \text{otherwise} \end{cases}$$

$W$  passes Feferman's criterion as presented above, but does not pass the modified criterion, to be presented in the next section.

The following operations are *not* homomorphism invariant:

- Identity, as a type  $(0^2 \rightarrow b)$  operation (see also Section 4.3.4).
- Cardinality quantifiers, that is:  $\exists_\kappa$  for  $\kappa \geq 2$ , as operations of type  $((0 \rightarrow b) \rightarrow b)$ , defined by:

$$\exists_{\kappa}^M(p) = \begin{cases} T & \text{if there are at least } \kappa \text{ distinct } x' \text{'s in } M_0 \text{ such that } p(x) = T \\ F & \text{otherwise} \end{cases}$$

- Second-order universal function quantification,  ${}^2\forall$ , as a type  $((0 \rightarrow b) \rightarrow b) \rightarrow b$  operation, defined by:

$${}^2\forall^M(f) = \begin{cases} T & \text{if } (\forall q \in M_{0 \rightarrow b}) f(q) = T \\ F & \text{otherwise} \end{cases}$$

(Feferman, 1999, pp. 40f)

We might add that the quantifier *most* taken either as a one place second-level predicate of type  $((0 \rightarrow b) \rightarrow b)$  or as a two place second-level predicate of type  $((0 \rightarrow b), (0 \rightarrow b) \rightarrow b)$ , is *not* homomorphism invariant. It may be claimed that leaving *most* out has a hint of arbitrariness. It is not completely clear why *all* should be on one side of the boundary and *most* on the other. (See a thorough discussion of the subject in (Sher, 1991)).

By being more restrictive, Feferman’s criterion solves the problems diagnosed with the Tarski-Sher thesis. It is more set-theoretically robust. It disallows second-order quantification and cardinality quantifiers. By looking at pairs of models of different sizes, it disallows the so-called “domain sensitive quantifiers”, and gives substance to the requirement of being the “same operation” across domains.

#### 4.3.4 The modified criterion and definability in FOL

Further support for taking homomorphism invariance as a criterion for logicity comes from Feferman’s main result.

**Definition 1** ( $\lambda$ -definability) An operation  $O$  is said to be  $\lambda$ -definable from operations  $O_1, \dots, O_k$  if it is given by a definition from them in the  $\lambda$ -calculus uniformly over each  $M$ , that is, if there is a term  $t(z_1, \dots, z_k)$  of the typed  $\lambda$ -calculus with constants  $T$  and  $F$ , where

each  $z_i$  is of the same type as  $O_i$  and  $t$  is of the same type as  $O$ , such that in each functional type structure  $M$ ,  $O^M = t(O_1^M, \dots, O_k^M)$  (Feferman, 1999, p. 41).<sup>7</sup>

**Definition 2** (Monadic types) A type  $\tau = (\tau_1, \dots, \tau_n \rightarrow \sigma)$  is said to be *monadic* if  $\sigma = b$ , and each  $\tau_i$  is either 0,  $b$  or  $(0 \rightarrow b)$  (see *ibid*, p. 42).

Restricting the scope to monadic types, we receive the following connection between homomorphism invariance and first-order logic:

**Theorem 2** (Feferman 1999, p. 43) *The operations  $\lambda$ -definable from the operations of the predicate calculus (FOL) without identity are exactly those definable from homomorphism invariant operations of monadic type.*

Feferman does not state explicitly what exactly he takes from his main theorem. However, it is obvious that it has philosophical implications analogous to those that McGee drew from his definability theorem: inasmuch as what ever is definable from FOL is intuitively logical, so homomorphism invariance should be taken as a sufficient condition for logicality. Unfortunately, it remains unclear why homomorphism invariance should be taken as a necessary condition on logicality: it rules out many counterexamples to the Tarski-Sher thesis, but there is no argument showing that it does not rule out too much.

Now, the theorem leads us to modify Feferman's criterion to a restricted form. According to the restricted version, only operations of monadic type that are homomorphism invariant are logical. This restriction would rule out the *wellfoundedness quantifier* that the original criterion allowed. In support of this restriction, Feferman cites works on natural language semantics that give monadic quantification a primary role (Keenan & Westerstahl,

---

<sup>7</sup>There is a slight difference between Feferman's terminology and our own: we have referred to  $M$  as *model* (a pair of a domain and an interpretation function), and Feferman as a *type structure*, but this makes no essential difference to the issues at hand.

1997).<sup>8,9</sup>

### 4.3.5 The problematic use of formalisms in the philosophical debate on logicity

At this point I would like to draw attention to issues arising from the work of Feferman and others on logicity that have to do with the particular way they use formal methods. The discussion is much inspired by (Casanovas, 2007), though the specific examples are not taken from that article.

As we have seen, both McGee (1996) and Feferman (1999) employ some definability results in support of their proposed criteria for logicity. McGee brought together isomorphism invariance and definability in  $L_{\infty\infty}$ , and Feferman connected homomorphism invariance with  $\lambda$ -definability in FOL without identity. Both of them used a particular background structure to frame their results. For instance, McGee looked at operations as working on variable assignments. Feferman works in a functional finite type framework, where all functions are considered that belong to the type structure built on the domain of individuals and the domain of truth values. Both McGee and Feferman receive desired results, at least to some extent: capturing the basic intuitively logical operations, so that standard logic passes their tests, and leaving out some intuitively nonlogical operations. But what they neglect to mention is how much depends on their specific choice of framework. Focusing on Feferman, we can take two main examples to deliver the point. The first is the truth-functional connectives. Feferman construes the truth functions as working on purely boolean types, thus restricting them to closed formulas. In a more general setting, where truth functional connectives apply to sets of variable assignments, they would not

---

<sup>8</sup>This argument is criticized by Bonnay (2008). Bonnay claims that the evidence from natural language does not fully support Feferman's argument, and casts doubt on the relevancy of natural language phenomena to the debate on logicity (Bonnay, 2008, p. 44).

<sup>9</sup>For some further discussion of the merits of monadic quantification see (Tharp, 1975, pp. 13ff).

pass Feferman's test. The second example regards identity and the functional type framework. Construed as a function, as in Feferman's framework, identity is not homomorphism invariant, but construed as a relation, it is (though non-identity isn't).

Starting with the first example, we look at the truth-functional connectives. Feferman takes the truth-functional connectives to denote functions of purely boolean types, that is, from sequences of truth values to truth values. The domain of individuals plays no role. Conjunction is simply the function that takes the pair  $\langle T, T \rangle$  to  $T$  and all other boolean pairs to  $F$ . All such functions are trivially homomorphism invariant. That is because the homomorphisms Feferman considers are all constant on  $M_b$ , i.e. on the truth values.

Now recall that McGee assimilates the truth-functional connectives to basic set-theoretic operations. Conjunction, on his construal, is assimilated to intersection. This makes sense when we look at conjunction as operating on sets of variable assignments: the set of variable assignments satisfying  $\phi \wedge \psi$  is the intersection of the set of variable assignments satisfying  $\phi$  and the set of variable assignments satisfying  $\psi$ .<sup>10</sup> Conjunction, in this setting, is not homomorphism invariant. Take the the following simple case, where we look at only the assignment for the variable  $x$ , where  $|M|$  is the domain of  $M$ :

$$\begin{aligned} |M| &= \{0, 1\}, \phi^M(x) = \{\langle 0 \rangle\}, \psi^M(x) = \{\langle 1 \rangle\} \\ |M'| &= \{a\}, \phi^{M'}(x) = \{\langle a \rangle\}, \psi^{M'}(x) = \{\langle a \rangle\} \end{aligned}$$

Let  $h$  be the only function from  $M$  to  $M'$ , which takes both 0 and 1 to  $a$ .  $h$  is surjective, that is, a homomorphism in Feferman's sense. Yet conjunction, construed as intersection, is not invariant under  $h$ :  $h(\wedge^M(\phi^M, \psi^M)) = h(\emptyset) = \emptyset$  while  $\wedge^{M'}(h(\phi^M), h(\psi^M)) = \wedge^{M'}(\phi^{M'}, \psi^{M'}) = \{\langle a \rangle\}$ . (It is noteworthy that disjunction, construed as union of sets of variable assignments, does pass the test of homomorphism invariance.)

Feferman's connectives are limited to the boolean realm, and thus apply only to closed

---

<sup>10</sup>As we pointed out in the preceding section, McGee ends up defining operations in a uniform way as functions from sequences of sets of variable assignments to sets of variable assignments, so conjunction is eventually not assimilated to intersection *per se*, but the point made here hold either way.

## CHAPTER 4. VARIETIES OF INVARIANCE

formulas—but the logic they are part of need not suffer from restricted expressibility. Employing the  $\lambda$  operator in addition to the connectives and quantifiers as Feferman sets them up, allows the full expressive power of standard first-order logic. Thus, I am not suggesting that Feferman’s choice of framework is in any sense flawed, just that there are alternative frameworks that yield different results.

Moving on to the second example, we note that in this setting, identity is construed as a *function*: the operation considered is the characteristic function of the relation of identity, of type  $(0^2 \rightarrow b)$ :

$$I^M(x, y) = \begin{cases} T & \text{if } x = y \\ F & \text{otherwise} \end{cases}$$

Casanovas (2007) points out that the use of the framework of functional types, as opposed to relational types, makes a difference. We can verify this claim with the case of identity. Identity is not preserved by homomorphisms that are not injective. However, note that if we were in a relational type framework, and construed identity as the relation  $I^M = \{ \langle a, a \rangle : a \in M_0 \}$ , it would turn out to be homomorphism invariant. For any homomorphism  $h : M_0 \rightarrow M'_0$ , extended to all relational types over  $M$  and  $M'$ ,  $h(I^M) = h(\{ \langle a, a \rangle : a \in M_0 \}) = \{ \langle h(a), h(a) \rangle : a \in M_0 \} = \{ \langle a', a' \rangle : a' \in M'_0 \} = I^{M'}$  (the penultimate equality holds because  $h$  is surjective).

Some might say that identity is a relation in its essence, not a function, and should be considered as such for any test for logicity. I make no such claim. The way identity is formalized is relative to a given framework. Surely the characteristic function of identity is a viable formalization as much as the underlying relation. We may say that there are different acceptable ways to represent identity, and the specific choice is an *artefact* of the system we choose. But then one should be cautious with the conclusions they draw from a given formalization, that do not follow in all the acceptable alternatives. In a functional type framework identity is homomorphism invariant and in a relational type framework



it isn't. In a boolean type framework conjunction is homomorphism invariant, and set as intersection it isn't.

Consequently, homomorphism invariance can serve as a criterion for logicity only relative to a specified framework. We should thus modify Feferman's criterion to be: *an operation in the functional finite type framework is logical if and only if it is a (monadic) homomorphism invariant operation* ("monadic" should be added in for the restricted form of the criterion). Feferman's criterion is not generally applicable, and this serves as a challenge to its philosophical significance.

The issue I raise here pertains to all formal accounts of logical terms, so it applies to the Tarski-Sher thesis as well as to Feferman's criterion. But in Feferman's case the matter is accentuated. The reason is that the notion of homomorphism is less robust than that of isomorphism. While the criterion of bijection invariance is insensitive to whether we have a relational or functional type structure framework, the criterion of homomorphism invariance is sensitive to the difference.

Does the relative nature of Feferman's criterion infringe its validity? It would appear that it doesn't. All logic is done in some framework or other. For instance, this whole work assumes a set theory as a basic framework, as opposed to a category theory or proof theory. It is then reasonable that philosophical considerations would be relative to a framework as well. In Feferman's framework, his criterion achieves all that he claims it does: it preserves basic standard logic and leaves out non-set theoretically robust and domain sensitive operations. The definability result, connecting his criterion with first-order logic gives further substance to his position.

Nonetheless, the choice of a functional type framework is on a different level from the choice of set theory as a background theory. This calls for extra precaution when evaluating or using his criterion. His specific choice of framework is not the most common or natural one, at least when dealing with philosophical questions of logicity. In other frameworks, homomorphism would not be the right relation to look at to achieve Feferman's results.

## CHAPTER 4. VARIETIES OF INVARIANCE

So one should not look at the nature of homomorphisms to gain better understanding of logicality.

Furthermore, Feferman does not provide us a tool to assess the logicality of various notions irrespective of their background framework. The overall features of the system Feferman is proposing, in his given framework, should be looked at. In this regard, Feferman's claims might be misleading.

Feferman distinguishes himself from holistic approaches to the question of logicality. By *holistic approaches* Feferman means “model-theoretic characterizations of logics as a whole, without attempting to isolate the separate contributions of individual operations which generate them” (Feferman, 1999, p. 48). Indeed, Feferman shows us that first-order logic without identity can be generated from a specific group of terms. Feferman indicates the merits of the terms of this group. But in essence, Feferman's approach has some holistic features to it: only in the context of a specific framework the group of terms enjoys its special qualities. It would thus be safer to say that Feferman does not propose a criterion for logical operations, but rather a whole *system* in which the boundaries of logic can be accounted for.

### 4.3.6 Concluding remarks

Feferman claims to have improved on the Tarski-Sher thesis. Indeed, the criterion for logicality that he proposes, in its given framework, satisfies the desiderata laid out for it. We observed that Feferman's criterion works only in a very specific framework. This affects the philosophical significance of Feferman's criterion. Rather than informing us about the nature of logical notions in general, due to its restricted applicability, Feferman's criterion how the boundaries of logic can be drawn in his specific framework.

We remain with one unresolved issue. The arguments Feferman employs are all in support of homomorphism invariance being a *sufficient* criterion for logicality. It rules out

## CHAPTER 4. VARIETIES OF INVARIANCE

the known problematic operations, and the definability result presumably shows that it keeps logic safely within its bounds. However, no argument shows that homomorphism invariance is a *necessary* condition. This matter is emphasized considering that there are operations that do not pass Feferman's criterion that can naturally thought of as logical, such as identity and *most*. In the next section we look into Bonnay's criterion, which, while being more restrictive than the Tarski-Sher thesis, does not impose homomorphism invariance as a necessary condition for logicality.

## 4.4 Bonnay: Potential Isomorphisms

### 4.4.1 Summary

The last position I discuss in this chapter is that of Denis Bonnay (2008). Bonnay offers his own modification to the Tarski-Sher thesis. By his account, an operation is logical if and only if it is invariant under potential isomorphisms. Any two isomorphic models are potentially isomorphic, and unlike isomorphism, a potential isomorphism can hold between domains of different sizes. This makes Bonnay's criterion more restrictive than the Tarski-Sher thesis.

By the nature of this survey, we cannot go into all the details of the accounts we deal with. Bonnay's description of his view is particularly rich with both mathematical and philosophical content, most of which will unfortunately have to be neglected in this discussion.

### 4.4.2 Bonnay's criterion

Bonnay concurs with the critics of the Tarski-Sher thesis, Feferman among them, that the criterion of isomorphism invariance exceedingly overgenerates. Bonnay also disagrees with Feferman's solution, since even though it solves some of the problems of the Tarski-Sher thesis, it is not justified as both a necessary and a sufficient condition on logicity. Keeping the idea of logical terms being invariant under some kind of relation between structures, Bonnay proposes a criterion that employs the notion of *potential isomorphism*. Let us begin with Bonnay's criterion itself and the definitions it requires. In the subsequent sections we look into Bonnay's arguments for his proposed criterion.

To make things easier, we look at the various operations we are interested in as classes of structures. For example, let  $Q$  be some, perhaps generalized, one place quantifier. The interpretation of  $Q$  in a model  $M$ ,  $Q^M$ , is a set of sets of individuals (or, interpreted functionally, a function from sets of individuals to truth values). But let us look at  $Q$  as

the class of structures of the form  $\mathcal{M} = \langle D, P \rangle$  such that  $P \subseteq D$  and for any model  $M$  with domain  $D$ ,  $P \in Q^M$  (or, in the functional form,  $Q^M(P) = T$ ). We write  $Q(\mathcal{M})$  to say that  $\mathcal{M}$  is a structure in the class  $Q$ .

In order to accommodate the truth-functional operators, we will assume that the truth values are part of the structures—they will be viewed as 0-place relations, and we assume that every isomorphism is constant on the truth values.

Also, we write  $|\mathcal{M}|$  for the first element in the structure  $|\mathcal{M}|$ , e.g. if  $\mathcal{M} = \langle D, P \rangle$ , then  $|\mathcal{M}| = D$ . In these terms, an operation  $O$  is isomorphism invariant if and only if for any two structures  $\mathcal{M}$  and  $\mathcal{M}'$ , if  $\mathcal{M}$  is isomorphic to  $\mathcal{M}'$  ( $\mathcal{M} \approx \mathcal{M}'$ ), then  $O(\mathcal{M})$  if and only if  $O(\mathcal{M}')$ .

Bonnay's criterion uses the notion of a *potential isomorphism*, which, in turn, is defined through the notion of *partial isomorphism*:

**Definition 3** (Partial isomorphism) Let  $\mathcal{M}$  and  $\mathcal{M}'$  be two structures, and  $f : |\mathcal{M}| \rightarrow |\mathcal{M}'|$  a function.  $f$  is a *partial isomorphism* between  $\mathcal{M}$  and  $\mathcal{M}'$  if and only if there are two substructures  $\mathcal{A}$  and  $\mathcal{A}'$  of  $\mathcal{M}$  and  $\mathcal{M}'$  such that  $f$  is an isomorphism between  $\mathcal{A}$  and  $\mathcal{A}'$ . (Bonnay, 2008, p. 46)

**Definition 4** (Potential isomorphism) A *potential isomorphism*  $I$  between two structures  $\mathcal{M}$  and  $\mathcal{M}'$  (notation  $I : \mathcal{M} \approx_p \mathcal{M}'$ ) is a nonempty set of partial isomorphisms such that:

for all  $f \in I$  and  $a \in |\mathcal{M}|$  (resp.  $b \in |\mathcal{M}'|$ ), there is a  $g \in I$  with  $f \subseteq g$  and  $a \in \text{dom}(g)$  (resp.  $b \in \text{rng}(g)$ ).

(Bonnay, 2008, p. 46)

Less formally: two structures are potentially isomorphic if there is an isomorphism between some substructures of them, and this isomorphism can be extended to include any individual in the domain. Potential isomorphisms allow relating models of different sizes, while maintaining a certain notion of sameness of structure. An operation  $O$  is invariant

## CHAPTER 4. VARIETIES OF INVARIANCE

under potential isomorphisms if for any potentially isomorphic structures  $\mathcal{M}$  and  $\mathcal{M}'$ ,  $O(\mathcal{M})$  if and only if  $O(\mathcal{M}')$ .

We shall refer to the following as *Bonnay's criterion*:

An operation is logical if and only if it is invariant under potential isomorphisms.

Let us look at some examples of operations that pass Bonnay's criterion.

- Negation, as a class of structures of the form  $\langle M, F \rangle$ .
- Conjunction, as a class of structures of the form  $\langle M, T, T \rangle$ .
- Disjunction, as a class of structures of the form  $\langle M, T, T \rangle$ ,  $\langle M, T, F \rangle$  or  $\langle M, F, T \rangle$ .
- Identity, as a class of structures of the form  $\langle M, a, a \rangle$  where  $a$  is an individual.
- Finite cardinality existential quantifiers,  $\exists_n$ , as classes of structures of the form  $\langle M, P \rangle$ , where  $\text{card}(P) = n$ .
- The quantifier *infinitely many*,  $\exists_{\geq \aleph_0}$  for  $n \in \omega$ , as a class of structures of the form  $\langle M, P \rangle$ , where  $\text{card}(P) \geq \aleph_0$ .

The following operations do *not* satisfy Bonnay's criterion:

- The cardinality quantifiers  $\exists_{! \kappa}$ , for  $\kappa \geq \aleph_0$ , as classes of structures of the form  $\langle M, P \rangle$ , where  $\text{card}(P) = \kappa$ .
- The cardinality quantifiers  $\exists_{> \kappa}$ , for  $\kappa \geq \aleph_0$ , as classes of structures of the form  $\langle M, P \rangle$ , where  $\text{card}(P) > \kappa$ .
- The cardinality quantifiers  $\exists_{\geq \kappa}$ , for  $\kappa > \aleph_0$  as classes of structures of the form  $\langle M, P \rangle$ , where  $\text{card}(P) > \kappa$ .

In addition, domain sensitive quantifiers do not pass Bonnay’s criterion, and the 2-place quantifier *most* passes it when we restrict ourselves to countable domains. Second-order logic is excluded by Bonnay’s criterion—it can express cardinality properties which are not invariant under potential isomorphisms—though, as Bonnay submits, a systematic development of the issue of logicity and invariance to higher-order operations is still missing (Bonnay, 2008, p. 64, f.n. 27).

Bonnay sums up the picture of logic that follows from his criterion:

- (i) The issue of the logicity of cardinality quantifiers is addressed in a selective way: only quantifiers which distinguish among infinite cardinals pass the test.
- (ii) All arithmetical truths are logical truths.
- (iii) Not all mathematical truths are logical truths.
- (iv) Second-order logic is not genuine logic.

(Bonnay, 2008, p. 64)

Bonnay’s criterion, as Bonnay points out, sets the limits of logic “somewhere between arithmetic and set theory” (*ibid*). He explains that number words are regular words and are part of natural language, so should reasonably be included in logic. And so is the difference between the finite and the infinite. Infinite cardinalities, says Bonnay, are not part of competent speech. Now of course the support Bonnay’s criterion can gain from these explanations of natural language is limited. After all, most of the vocabulary of natural language does not pass Bonnay’s criterion when formalized. Simple analytic entailments (Tom is a bachelor; therefore Tom is unmarried) are not included in Bonnay’s logic, even though they are as much a part of natural language as basic arithmetic. But Bonnay offers further arguments in favor of his criterion—those are discussed in the next two sections.

### 4.4.3 Generality

Logic is said to be the most general discipline, and some incorporate this feature in the criteria they propose for logical terms or operations. Tarski, for instance, indicates that his criterion of permutation invariance yields notions of a very general character (Tarski, 1986, p. 149). Recall that Tarski considered various disciplines as characterized by invariance conditions in Klein’s Erlangen Program. Geometric disciplines are characterized there through the classes of transformations of the domain under which their notions are invariant. The widest class of transformations, the class of all permutations of a domain, would yield the most restrictive class of notions, and presumably those with the most general nature.

As Bonnay points out, the generality received when looking at all transformations of a fixed domain is limited. Allowing the domain to vary would increase the generality of the operations received. Once we start looking at varying domains, instead of one fixed domain, we need to divert our attention from *transformations of the domain* to *functions or relations between structures sharing a signature*. The widest relation is then simply the universal one, holding between any two structures that have the same signature. However, the universal relation will not give us much of a logic: it allows only operations that include all or none of the structures of a given form. Thus, some limitations should be imposed on the scope of the relations involved.

Bonnay proposes the following qualifications to the principle that logic should be maximally general. First, he requires that the truth-functions, functional application and first-order functional application be considered as logical. This requirement already puts the universal relation out of the picture.

Secondly, Bonnay requires the following principle:

PRINCIPLE OF CLOSURE UNDER DEFINABILITY. An interpreted symbol definable only by means of logical constants is a logical constant.

(Bonnay, 2008, p. 50)



Making the principle of definability mathematically precise, Bonnay requires that for any purely logical system  $L$  and for any operation  $O$ , if there is a sentence  $\phi_O$  of  $L(\bar{R})$  such that:

$$O(\langle M, R \rangle) \text{ iff } \langle M, R \rangle \models_L \phi_O$$

then  $O$  should be considered as logical. Bonnay contends that “Operators which are definable in a purely logical manner are logical. We just do not see how a non-logical element could creep in the logical elements of the definition and make the defined operator non logical” (*ibid*).

Is Bonnay right in his contention? Leaving it to mere intuitions, it could really go either way. Here is a suggestion for something that might “creep in” and make an operation definable from logical operations nonlogical. If, for instance, logical operations are viewed by us as basic building blocks of logic, we might require them to be *simple* in some sense. Allowing closure under definability would bring in operations that might be defined in very complicated ways from the basic ones. Thus, whereas we might view the standard quantifiers as quite basic and simple, a quantifier that is defined by a string of multiple quantifiers might not seem so simple.

However, if we don’t have this requirement of simplicity, Bonnay’s contention is more than plausible. Dropping simplicity, why should a construction of logical operations not maintain the logical nature of the particles? But on the other hand, shouldn’t we first explicate the notion of logicity and then see whether the principle of closure under definability come out as a result? In many accounts of logicity it does. McGee and Feferman’s definability results show this with respect to Tarski’s thesis and Feferman’s criterion.<sup>11</sup> But Bonnay imposes closure under definability as a *condition* on logical terms: it serves for

---

<sup>11</sup>Closure under definability is a direct result in criteria such as the one presented in (Carnap, 1937) by which logical expressions are those that are constructed from a special class of expressions  $\mathcal{R}$  such that any sentence constructed just by expressions from  $\mathcal{R}$  is determinate.

## CHAPTER 4. VARIETIES OF INVARIANCE

him as an assumption rather than a result of logicity. So presumably, when spelling out Bonnay's view, we should point out that closure under definability is a *feature of logicity* that should be taken into account when spelling out a criterion.

Now, if generality would have us relate any two structures sharing a signature, Bonnay's restricted form of generality would have look at the most comprehensive such relation  $S$  such that the truth functions, functional application and the existential quantifier are  $S$ -invariant, and such that the  $S$  invariant notions are closed under definability. Bonnay shows that this relation is potential isomorphism, and the generality of logic provides an argument in favor of his criterion.

### 4.4.4 Formality

Logic is also said to be *formal*. Bonnay considers the view by which formal notions are those which are insensitive to arbitrary switchings of objects (p. 34). This characterization of formality leads to Tarski's thesis according to which logical notions are invariant under permutations. The problem with this characterization is that it does not distinguish between logic and other mathematical disciplines such as set theory, and so highly overgenerates. Bonnay thus seeks a way to strengthen the requirement of formality so as to keep logic in its own bounds.

Bonnay relates formality to *lack of content*. *Formal* notions are those that lack any content that is sensitive to the identity of objects. *Logical* notions are, according to Bonnay, also free of *problematic set-theoretic content*. The notions that are free from such content, according to Bonnay, are those that are set-theoretically *absolute*. Bonnay endorses Feferman's requirement of absoluteness with respect to ZFC, which put imprecisely says that the meaning of a logical operation is the same in all models of set theory (see (Bonnay, 2008, p. 56)).

Bonnay thus requires that logical notions be both deprived of non-formal content and

of problematic set-theoretic content (p. 60). Invariance under isomorphisms ensures the first requirement, and adding on absoluteness gives an appropriate bound between logic and mathematics. Bonnay shows that if those conditions are taken as both necessary and sufficient, the logical notions are those that are invariant under potential isomorphisms. Formality, in its strengthened form, gives a further argument for Bonnay's criterion.

#### 4.4.5 Concluding remarks

The view that logic is both formal and general can easily lead to the Tarski-Sher thesis on logical terms (see (Sher, 1996, p. 674)). However, since the Tarski-Sher thesis does not make a proper distinction between logic and set theory or mathematics in general, Bonnay suggests to impose some further constraints. Regarding logical notions as those that are the most general ones that are closed under definability and include the basic notions of standard first-order logic yields Bonnay's criterion. And so does regarding logical notions as those that are formal in the sense of insensitivity to the identity of objects and are also absolute with respect to ZFC. Bonnay thus has two independent arguments to the effect that logical notions are those that are invariant under potential isomorphisms. We pointed out that closure under definability is a result of the criteria for logicality discussed in previous sections, and that Bonnay takes it as an *assumption*. This makes closure under definability a basic feature of logicality in Bonnay's view, and excludes, for instance, a view by which logical notions are simple.

## Chapter 5

# Epistemic Criteria for Logical Terms

### 5.1 The Epistemic Component in Logical Consequence

In Chapter 3 I formulated *Necessity* and *Formality* as two basic conditions on logical consequence. It may seem that I left out an epistemic component which is as basic and important as those two conditions. Accepting that there is such a component, I have two reasons for leaving it out of the basic conditions. The first is that I do not wish to limit the discussion to a specific formulation of an epistemic condition, when there are very different ways in which the epistemic component can be understood. The second reason is that I see an epistemic motivation as *underlying* the two conditions of *Necessity* and *Formality*. *Necessity* requires that logically valid arguments be truth-preserving across all possible worlds. This means that we don't have to know in which possible world we reside to draw a logically valid inference. *Necessity* thus makes logical practice independent of knowledge of contingent facts. *Formality* can be epistemically motivated in different ways. In Chapter 3 I suggested that formality relieves us from having to answer intricate metaphysical questions in order

to argue validly. *Formality* thus makes logical practice independent of knowledge of some necessary facts.<sup>1</sup>

Nonetheless, many philosophers accept the following epistemic condition:

(*Apriority*) Logical consequence is knowable a priori.<sup>2</sup>

The “epistemic approaches” I discuss in this section use epistemic notions in demarcating logical terms. The idea is that logical consequence and logical truth are knowable a priori, and that this requires of the logical terms to be “a priori” in a sense to be specified. In Sections 5.2- 5.3 I discuss Peacocke’s criterion for logical terms. In Section 5.4 I briefly present McCarthy’s modification to Peacocke’s criterion, which is based on a combination of an “invariance criterion” with modal and epistemic notions. I discuss McCarthy’s objection to unmodified invariance criteria and claim that the objection is misguided.

## 5.2 Peacocke’s Criterion for Logical Terms

Peacocke’s criterion, presented in his 1976 paper, is roughly this:

$\alpha$  is a logical constant if  $\alpha$  is non-complex and, for any expressions  $\beta_1, \dots, \beta_n$  on which  $\alpha$  operates to form expression  $\alpha(\beta_1, \dots, \beta_n)$ , given knowledge of which sequences satisfy each of  $\beta_1, \dots, \beta_n$ , and of the satisfaction condition of expressions of the form  $\alpha(\gamma_1, \dots, \gamma_n)$ , one can know a priori which sequences satisfy  $\alpha(\beta_1, \dots, \beta_n)$ , in particular without knowing the properties and relations of the objects in the sequences. (Peacocke, 1976, p. 223)

---

<sup>1</sup>The epistemic motivation I allude to can be explained by the requirement that logical practice would be both practical and safe: we may be ignorant of various facts and still be assured that we can get from true premisses to a true conclusion. This is by contrast to a different, possible, epistemic approach, requiring the *completeness* of logical consequence: that every logically valid argument be provable. I take this last approach to describe a slightly different notion of logical consequence than the one characterized by *Necessity* and *Formality*, similarly to (Shapiro, 1998, 2002, 2005).

<sup>2</sup>Some would add the reservation: “...when it is knowable” (see (Hanson, 1997, p. 378).

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

The idea is: you may need all sorts of knowledge to determine the truth value or satisfiability of a formula  $\alpha(\beta_1, \dots, \beta_n)$  containing a logical constant  $\alpha$ , but the application of  $\alpha$  itself requires only a priori means. The knowledge needed for determining the semantic value of  $\alpha(\beta_1, \dots, \beta_n)$  may include only knowledge of the semantic values of  $\beta_1, \dots, \beta_n$  and a priori knowledge. Peacocke does not confine logicity to formula-forming operators, and proposes an analogous criterion for functions. This formulation is only supposed to give the basis for a criterion, since many modifications are needed to make it work.

Now let us see how Peacocke's rough characterization fares with the standard quantifiers and connectives, and the required modifications and additional assumptions will stem from there. The case of negation is the simplest. If one knows whether a sequence  $s$  satisfies  $\beta$ , then, on the condition that they know the satisfaction rule for negation, they will be in a position to know a priori whether  $s$  satisfies  $\neg\beta$ . So negation satisfies Peacocke's criterion straightforwardly (putting aside qualms about the notion of the *a priori* itself and coherence of the criterion, to be discussed later). But negation seems to be the only case that can be handled without a problem. Already when we get to conjunction things look differently, as we shall immediately see. Interestingly, Peacocke shows that quantification is problematic via an argument that also shows directly that conjunction is problematic, but he does not discuss conjunction. He even presumes that sentential connectives in general require no modification of his criterion: "for a sentential connective  $\alpha$ , one can know whether  $s$  satisfies  $\alpha(\beta_1, \dots, \beta_n)$  given only knowledge of the satisfaction clause for  $\alpha$  and of whether  $s$  itself satisfies  $\beta_1, \dots, \beta_n$ ." (Peacocke, 1976, p. 223).

Later on, when considering the existential quantifier, Peacocke notes that the following inference is invalid:

$x$  knows of  $y$  that it's  $F$  and  $x$  knows of  $y$  that it's  $G$

So,  $x$  knows of  $y$  that it's  $F$  and  $G$

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

since  $x$  may not realize that it is the same object that is both  $F$  and  $G$  (Peacocke, 1976, p. 227).<sup>3</sup> Clearly, this argument refutes Peacocke's claim about the connectives, where  $n = 2$ ,  $\beta_1 = F$ ,  $\beta_2 = G$ ,  $s = y$  and  $\alpha$  is conjunction.<sup>4</sup> According to Peacocke's criterion, if conjunction is a logical constant, then given knowledge of which sequences satisfy  $F$  and which sequences satisfy  $G$ , and of the satisfaction condition of conjunction, one should be able to know a priori which sequences satisfy  $F \& G$ . That is, it seems like Peacocke simply needs to allow knowledge that  $y$  is  $F$  and  $G$  on the condition of knowledge that  $y$  is  $F$  and that  $y$  is  $G$  if conjunction is to be considered logical, although he admits that this inference is invalid.

Moving on to the existential quantifier, Peacocke notes that since the previous inference is invalid, so is this one:

$x$  knows that  $s' \approx_i s$  and knows that  $s'$  satisfies  $A$  (where ' $s$ ', ' $s'$ ' are variables)

So,  $x$  knows that:  $s' \approx_i s$  and  $s'$  satisfies  $A$

where  $s' \approx_i s$  means that the sequences  $s'$  and  $s$  agree on the  $i$ th component (Peacocke, 1976, p. 227). This kind of inference is required for evaluating whether the satisfaction conditions for the existential quantifier are met. In order to evaluate whether  $s$  satisfies  $\exists x_j A$ , one needs to consider all sequences  $s'$  that differ from  $s$  in at most the  $j$ th place, that is agree with  $s$  on all  $i \neq j$ . So knowledge that two sequences agree on some element needs to be assumed. With similar considerations, Peacocke shows that knowledge that two sequences differ on some element needs to be assumed as well.

---

<sup>3</sup>See also (McCarthy, 1981, p. 503) and (Quine, 1956).

<sup>4</sup>One might point to the difference between sentential conjunction and the conjunction of predicates, which we seem to have in this example. I take it that the difference between ' $x$  knows of  $y$  that it's  $F$  and  $G$ ' and ' $x$  knows that:  $y$  is  $F$  and  $y$  is  $G$ ' as insubstantial. I presume that Peacocke as well does not consider the difference to be substantial, given that he uses the invalidity of the inference using the conjunction of predicates to argue for the invalidity of an inference using sentential conjunction, to be quoted below.

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

The quantifiers impose yet another requirement. Take for instance the satisfiability condition for the universal quantifier:  $s$  satisfies  $\forall x_i A$  iff all sequences  $s'$  differing from  $s$  at most at the  $i$ th component satisfy  $A$ . In order for one to know a priori which sequences satisfy  $\forall x_i A$  it does not suffice to know which sequences satisfy  $A$  and the satisfiability condition for  $\forall$ : one also has to know that the sequences they know of are all the sequences there are (see (Gómez-Torrente, 2002, p. 23) and (Peacocke, 1976, p. 228)).

Peacocke also considers operators of intensional logic, such as ‘in the past’ and ‘necessarily’. Peacocke notes that similar amendments as those needed for the quantifiers may be posed for these operators. That is, the quantifiers require us to look at the totality of all sequences, why not allow also knowledge of satisfaction by sequences at all times or states of affairs? Indeed, Peacocke leans towards accepting the tense operator as logical on the basis of the analogy with the quantifiers (p. 224). And on the assumption that ‘necessarily’ is also interpreted analogously (looking at states of affairs instead of points in time), it should be allowed as well (p. 236). As for set-membership, Peacocke makes no allowance for this relation, contending that knowledge that a sequence  $s$  satisfies  $x_i \in x_j$ , given knowledge of its  $i$ th and  $j$ th elements, is not a priori (p. 237).

Peacocke notes three main virtues of his criterion. The first is that expressions meeting it will be topic-neutral: the satisfaction of  $\alpha(\beta_1, \dots, \beta_n)$  depends on no particular features, objects, properties or relations on which the satisfaction of  $\beta_1, \dots, \beta_n$  does not already depend (Peacocke, 1976, p. 229). The second is that the notion of logical consequence naturally generated by his criterion is the standard model-theoretic notion of validity, rather than the “substitutional” one<sup>5</sup> (*ibid* p. 230). The third is that the criterion may explain the plausi-

---

<sup>5</sup>By the “substitutional approach” to logical consequence, an argument is valid if it is (necessarily) truth-preserving under all uniform substitutions of the nonlogical terms with terms of the same syntactic category. The substitutional approach yields the arguably undesirable result where the logical validity of an argument depends on the richness of the substitution classes of its terms, and thus on the richness of the language. Tarski’s semantic approach, through the model-theoretic definition of logical consequence, blocks



bility of the inferential criterion for logical terms, that a term is logical iff its introduction and elimination rules wholly determine its meaning (p. 230f).

### 5.3 Criticisms of Peacocke's Criterion

Except for negation, all the standard logical terms of first order logic require some modification of Peacocke's basic criterion.<sup>6</sup> McCarthy, who offers an improvement on Peacocke's criterion (see Section 5.4) rightly senses "a trace of the ad hoc" in these modifications (McCarthy, 1981, p. 504). Peacocke himself submits that he has not provided a sharp boundary between logical and nonlogical terms, and yet he has given a principled characterization (Peacocke, 1976, p. 235). But the problem with the criterion is not its vagueness. The problem is that it doesn't fulfill the role of a criterion *at all*. One would expect that a criterion to be able to tell the definite cases, if not the borderline ones. However, Peacocke does not give a condition that can be used to test a given term for its logicity, with some borderline cases. Since most of what are supposed to be clear-cut cases (the standard logical terms) require special modification of the principle, it seems like if a boundary between the logical and the nonlogical exists, the principle is not a guide to where the boundary passes.

Peacocke's approach does not provide us a test for logical terms, but it does provide a thorough and useful analysis of the knowledge required by various terms. One can even put the epistemic talk aside, and then what Peacocke's analysis provides is the facts that determine for various  $\alpha$ , whether  $\alpha(\beta_1, \dots, \beta_n)$  is satisfied by some sequence  $s$  (besides those that determine the satisfaction of  $\beta_1, \dots, \beta_n$ ).

McCarthy accuses Peacocke's criterion of over-generating, e.g. in the case of some arithmetical terms. The term "successor" ( $Sc$ ), given the following rule, is an example:

---

this result (see (Tarski, 1936)).

<sup>6</sup>The problem we pointed out with regard to conjunction is shared by all binary connectives – nothing in the argument depended on the specific meaning of conjunction.

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

( $Sc$ )  $\forall s(s$  assigns to  $\ulcorner Sc(v) \urcorner$  the successor of  $s(v)$ , if  $s(v)$  is a natural number, and assigns  $s(v)$  to  $\ulcorner Sc(v) \urcorner$  otherwise). (McCarthy, 1981, p. 506)

As Peacocke, McCarthy holds that logical terms ought to be topic-neutral.  $Sc$ , according to McCarthy, is an “epistemically transparent expression”, i.e. it satisfies Peacocke’s criterion, yet it is intuitively topic-specific (*ibid*).

Peacocke has also been criticized for appealing to the notion of the *a priori*. Warmbrød, for instance, invokes Quinean skepticism about the *a priori* (Warmbrød, 1999, p. 506). Peacocke and Boghossian claim that what Quine refuted is that anything could be true by virtue of meaning, so his criticism of the notion of the *a priori* applies only inasmuch as the *a priori* is explained by the notion of *analyticity*. Rationalists who do not make this move, viewing the *a priori* as an epistemic notion that does not amount to truth in virtue of meaning, are not susceptible to these objections (Boghossian & Peacocke, 2000, p. 4).<sup>7</sup> Nevertheless, the notion of the *a priori* might still be said to be too problematic, and without a satisfactory theory of it being offered (in the case of Peacocke’s (1976), where he lays out his criterion), it cannot be helpful in the clarification of logicity or other concepts.

Another problem Warmbrød raises is that Peacocke (in (Peacocke, 1976)) does not explain the connection between his criterion for logical terms and logical truth or consequence. More specifically, why should the terms that pass Peacocke’s criterion be the ones that are *fixed* in a semantic definition of logical consequence? Why would fixing them render logical truth *a priori*? ((Warmbrød, 1999, p. 506), see also (Gómez-Torrente, 2002, p. 24)). Here we come back to a point raised in Chapter 2. Logical terms can be demarcated on the basis of the relation of logical consequence aimed to be captured, so that a proposed criterion for logical terms would be evaluated according to the consequence relation it yields. Alter-

---

<sup>7</sup>See also (Boghossian, 1997) where it is argued that Quine’s criticism of the *a priori* is valid when *a priori* is understood as *analytic* in one sense (as a metaphysical notion) and not in another (as an epistemic notion). However, we should note that the notion of the *a priori*, even if not understood through the notion of *analyticity*, is conflated by Quine’s holism (see (Sher, 2002)).

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

natively, one can formulate a criterion on the basis of direct intuitions or understanding of the concept of ‘logical-termhood’, leaving aside the connection to logical consequence (yet assuming these two concepts cohere).

Peacocke does not say explicitly which method he is using, or what he takes the significance of the distinction between logical and nonlogical terms to be. He simply proposes a criterion. However, Peacocke does seem to assume that logical terms can be characterized independently of logical truth or consequence (“But if ... features [differentiating logical constants] can be specified, why not use them directly in the criterion and cut out the detour through sentences or arguments?” p. 222). Thus, without further justification of his approach, it seems like Peacocke draws on an independent intuition or understanding of ‘logical-termhood’, regardless of the notion of logical consequence.

In (Peacocke, 2004) we get another picture of logical terms and their relation to logical truth and validity. In this later article, Peacocke lays out a realist account of logical terms, their understanding and role in inference. Peacocke characterizes there logical terms inferentially, through their natural deduction introduction and elimination rules. The rough characterization (which is not phrased as a strict criterion) is that the rules for logical terms are primitively obvious for anyone who understands the terms. Again, as in (Peacocke, 1976), some cases require modifications to the rough characterization, and these modifications pertain to a relation between thinkers and logical terms having to do with understanding those terms. With some additional principles and claims Peacocke sets forth, it follows that the rules for logical terms are truth-preserving. Validity is characterized by Peacocke as necessary truth-preservation under proper assignments to the logical terms (Peacocke, 2004, p. 182). So in Peacocke’s later approach, the concept of logical term is prior to that of validity. Logical terms may be constrained by principles that have to do with truth-preservation, but not by *logical validity* as such, which is defined on the basis of logical terms.

In (Peacocke, 2004) the a priori is no longer an assumed, unexplained notion. Peacocke

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

addresses the epistemic status of logical terms through *meaning-constitutive* inferences. Peacocke’s explanation there is nonetheless problematic, by Quinean as well as other considerations. Peacocke considers the introduction and elimination rules of conjunction:

$(\&I)$	$(\&E_1)$	$(\&E_2)$
A B	A&B	A&B
———	———	———
A&B	A	B

Peacocke then claims:

Anyone who understands ‘&’ finds instances of these principles compelling. That is, we can fix upon particular contents A, B and say: it strikes one who understands ‘&’ as obvious that given A together with B, then A&B. (Peacocke, 2004, p. 164)

The primitively compelling introduction and elimination rules constitute the meaning of conjunction, and inference based on them would be a priori. Peacocke considers conjunction as a relatively unproblematic case, while the rules for other terms, such as the existential quantifier, are less obvious and require a more complex treatment. But problems arise already with the simpler case of conjunction. Note that Peacocke does not specify the language to which ‘&’ belongs. Presumably, one can be said to understand a term only insofar as it is a term in a given language. Now Peacocke might have left out the phrase “in English” for simplicity’s sake. But in that case, why use the symbol ‘&’, and not just the English word ‘and’?

Another option is that ‘&’ should be considered as an idealized ‘and’, as the formal conjunction of propositional calculus. But in that case, the primitive obviousness of the rules for logical terms would not be a means to tell them apart from other terms: in standard propositional (or predicate) calculus nonlogical terms are not given any introduction and

elimination rules. If we suppose the language was modified so that some nonlogical term will have introduction and elimination rules, there is still no reason to think there would be an epistemic difference between those rules and the rules for conjunction. They are all stipulated and the question of their obviousness does not arise.

Perhaps ‘&’ really stands for the word ‘and’ in English (and perhaps for the synonymous words in other languages). But would the inference rules still be primitively obvious? According to some theories of natural language semantics and pragmatics, these rules might even be wrong; or, if they are correct as semantic rules, they defy pragmatic intuitions and are therefore not obvious,<sup>8</sup> see (Blakemore & Carston, 2005; Grice, 1989). In any case, it should be noted that Peacocke’s claim is empirical, and should be given empirical justification.<sup>9</sup>

## 5.4 The Modal and Epistemic Arguments against the Invariance Criterion and McCarthy’s Modification

McCarthy views Peacocke’s criterion as a necessary but insufficient criterion for logical terms. He supplements it by adding a requirement that logical terms be invariant under isomorphisms (for a discussion of invariance under isomorphism, see Chapters 3 and 4). McCarthy claims that invariance as well is an insufficient condition for logicity. In his (McCarthy, 1981) he proposes a criterion that combines Peacocke’s criterion and the invariance criterion to yield the desired result of logical terms that are both topic-neutral and a priori. In later writings ((McCarthy, 1987) and (McCarthy, 1998)), the connection to Peacocke’s criterion is looser, as the insufficiency of the invariance condition is diagnosed

---

<sup>8</sup>So, for instance, the commutativity of ‘and’ might be disputed if a temporal element is attributed to it, as in the examples: ‘She gave up philosophy and felt much happier’ and ‘She felt much happier and gave up philosophy’.

<sup>9</sup>See (Sher, 1999) for more criticism of characterizing logic through what is *obvious*.

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

by McCarthy as having to do with necessity as well as apriority.

The argument against the sufficiency of invariance was made by McCarthy in the presentation of his criterion, and was used by others in criticisms of Sher’s approach. The argument has a “modal” version and an “epistemic” version. In what follows I show that in both versions the argument is misguided, and stem from a confusion between object-language and metalanguage.

Let us first look at the simpler, modal, version (McCarthy, 1981, p. 514-515). Suppose that ‘ $K$ ’ is some contingently true statement of the metalanguage.<sup>10</sup> Consider a unary connective ‘ $N$ ’ in the object-language defined as:

$$(7) \forall s(s \text{ satisfies } \ulcorner N\phi \urcorner \leftrightarrow [(\sim s \text{ satisfies } \phi \wedge K) \vee (s \text{ satisfies } \phi \wedge \sim K)])$$

(McCarthy, 1981, p. 514)

Since  $K$  is true,  $N$  is coextensional with negation, and is invariant under isomorphism. If we allow  $N$  as a logical term, as Sher’s criterion does, the following sentence would be logically true, for any sentence  $\phi$ :

$$N\phi \leftrightarrow \sim \phi$$

But, McCarthy claims, in any counterfactual situation in which ‘ $K$ ’ is false, the sentence  $N\xi \leftrightarrow \sim \phi$  is false. But then “it is a reasonable constraint on any theory of validity for a language  $L$  that a sentence of  $L$  that it construes as a logical truth is *not* possibly false in  $L$ ” (*ibid* p. 515).

The epistemic version of the argument against the sufficiency of invariance under isomorphism is analogous to the modal version, where  $K$  is an a posteriori statement in the metalanguage. In that case, McCarthy claims that ‘ $N\psi \wedge \sim \psi$ ’ would be defeasible by

---

<sup>10</sup>McCarthy also assumes that  $K$  is a posteriori, but that assumption is redundant in this version of the argument.

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

reference to empirical information, though a logical truth by the invariance condition (*ibid*, p. 516).<sup>11</sup>

On the basis of these arguments, McCarthy formulates a criterion for logical terms that is supposed to take care of necessity and apriority as well as topic-neutrality. The basic criterion of invariance requires that logical terms be invariant under isomorphic models (see Section 4.1). McCarthy defines *rigid invariance* as invariance under isomorphic *possible situations* (McCarthy, 1987, p. 426), and *epistemic invariance* as invariance under isomorphic *epistemically possible situations* (*ibid*, p. 439). According to McCarthy, terms that are invariant under isomorphic models, rigidly invariant and epistemically invariant would yield the desired result of topic-neutrality, necessity and apriority (see (McCarthy, 1981, 1987, 1998)). I shall not go into the details of McCarthy's suggestion, but instead criticise his objection to the basic criterion of isomorphism invariance.

The problem with the modal argument is that it relies on an invalid inference from the modality of a sentence in the metalanguage to the modality of a sentence in the object-language. Take a look at the following inference:

- (1) It could have been that '2' referred to 3, while '+', '=' and '4' referred to +, = and 4 respectively.

---

<sup>11</sup>Hanson makes a similar argument against Sher's criterion in (Hanson, 1997), to which Sher responds in (Sher, 2001), replied by Hanson in (Hanson, 2002). Sher's ultimate reply to the modal argument is reference to her condition (B) in her criterion for logical terms ((Sher, 2001, p. 249), (Sher, 1991, pp. 56, 64)). According to (B), a logical constant is defined by a single extensional function, and is identified with its extension. In other words, logical terms are rigid, in the sense that their definitions in the metatheory are rigid. Sher does not pose apriority as a requirement on logical consequence. Due to its problematic nature, Sher wishes to remain neutral on the subject of the *a priori* (though she contends that posing a further condition excluding empirical terms does not conflict with her approach, only limits its generality). Hanson claims that given her assumptions, Sher cannot be neutral with respect to the apriority of logical consequence (Hanson, 1997, p. 392). Sher claims however that her account is in no way committed to the contentious concept of the *a priori* (Sher, 2001, pp. 257-260).

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

(2) So, ‘ $2+2=4$ ’ could have been false.

This, so far, is valid. (1) is true under minimal metaphysical assumptions: that symbols in arithmetic could have been used otherwise than they actually are. (2) is thus true since in a world where ‘2’ meant 3, ‘ $2+2=4$ ’ would have been a false sentence. This does not infringe its status as a necessary arithmetical truth in our language. It does not follow that in that world ‘ $2+2=4$ ’, as *we* mean it, wouldn’t have been true. So from (2) it does not follow that:

(3)  $\diamond \neg(2+2=4)$

The arithmetical case is analogous to McCarthy’s example. As it happens,  $K$  is true, so  $N$  refers to the functor of negation. Indeed, what  $N$  refers to is a contingent matter. But so is the meaning of ‘2’.

McCarthy tells us that in some world, where ‘ $K$ ’ is false, ‘ $N\phi \leftrightarrow \sim \phi$ ’ is false. Note that in that world ‘ $N\phi \leftrightarrow \sim \phi$ ’ is also *logically* false, while in ours it is logically true. They simply use a different language in that world. Logical truths should presumably be true in all possible worlds, but so long as their meanings are fixed. As vexed an issue as meaning is, that much is assumed when it is assumed that logical truths are necessary truths. That is, talk of necessary truth that employs possible worlds assumes that sentences can be evaluated in possible worlds *given the meanings that they have*.

McCarthy’s claim that the invariance criterion yields logical truths that defy the condition of necessity (that all logical truths are necessary) is thus falsified, for the intended meaning of that condition. But is ‘ $N\phi \leftrightarrow \sim \phi$ ’ a necessary truth? Rule (7) McCarthy gives us is presumably a model-theoretic rule, assigning an extension to  $N\phi$ , for all  $\phi$ , in each model. Assuming ‘ $K$ ’ is true, ‘ $N\phi \leftrightarrow \sim \phi$ ’ is true in all models. But is it true in all possible worlds?

In Chapter 3 I discussed Stewart Shapiro’s *blended approach*, by which models represent possible worlds under interpretations of the nonlogical vocabulary. This approach ensures



## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

that all logical truths are necessary truths, on the assumption that all possible worlds are represented by models. So, on that assumption, ‘ $N\phi \leftrightarrow \sim \phi$ ’ is necessarily true.

Can we avoid Shapiro’s assumption that all possible worlds are represented by models? This would require us to be able to assign truth-values to sentences in the object-language relative to possible worlds *directly*, without the mediation of models. Assuming we can do that, it is only object-language expressions that would be re-interpreted at different possible worlds. The metalanguage, just as in the model-theoretic interpretation, is used for telling us how to interpret the object-language, and is not re-interpreted or re-evaluated itself. Thus, observing rule (7), the only thing that can vary the truth-value of ‘ $N\phi$ ’ at different possible worlds is the truth-value of  $\phi$ . Thus, given that  $K$  is a true statement in the metalanguage, ‘ $N\phi \leftrightarrow \sim \phi$ ’ would be a necessary, logical truth.<sup>12</sup>

The epistemic version of the argument for the insufficiency of invariance under isomorphisms is more complicated, due to the notion of the *a priori* employed. The epistemic version has the same structure as the modal one, where  $K$  is an *a posteriori* truth, and the worry is that ‘ $N\phi \leftrightarrow \sim \phi$ ’ would be logically true yet *a posteriori*. If we were to deal with the apriority of ‘ $N\phi \leftrightarrow \sim \phi$ ’ as we did with its necessity, we would have to appeal to something like “epistemically possible worlds”. For anyone who equates *a priori* truth with truth at all “epistemically possible worlds”, our previous argument modified appropriately would suffice. However, this equation is more contentious than that of necessary truth with truth at all possible worlds. Arguably, the notion of the *a priori* does not pertain to some type of possibilities where a sentence is true, but to the way the truth of a sentence is determined, justified or known.

The question then is how an *a posteriori* statement of the metalanguage appearing in a

---

<sup>12</sup> One may object to my reliance on the requirement that  $K$  is in the metalanguage.  $K$  might also be in the object-language, as well as in the metalanguage. In that case,  $N$  should be seen as either an abbreviation or as mutually constraining the interpretation of  $K$ . But then  $N$  wouldn’t pass the test of invariance under isomorphism, as long as the interpretation of  $K$  is allowed to vary as well.

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

semantic rule of interpretation for a term in the object-language would affect the apriority of sentences containing that term. My first claim is that the answer to this question is not obvious. Any shared intuitions or defining features of the a priori pertain to natural language sentences, which are not primarily understood by us through rules of interpretation in a metalanguage. Now suppose that  $K$  is an a posteriori statement in the metalanguage, and the object-language term  $N$  is defined again by the rule:

$$(7) \forall s(s \text{ satisfies } \ulcorner N\phi \urcorner \leftrightarrow [(\sim s \text{ satisfies } \phi \wedge K) \vee (s \text{ satisfies } \phi \wedge \sim K)])$$

McCarthy contends that since our knowledge of the truth of the sentence ' $N\phi \leftrightarrow \sim \phi$ ' depends on the correct application of rule (7), and thus on knowledge of  $K$ , it is therefore a posteriori. However, it's not obvious that the a posteriori status of a statement in the metalanguage confers its status to sentences in the object-language. As pointed out with regard to the modal version of the criticism of the invariance condition,  $K$  is part of what determines the meaning of  $N$ . It's not the truth of ' $N\phi \leftrightarrow \sim \phi$ ' that depends on the a posteriori (or contingent) truth of  $K$ , it's the *meaning* of it that depends on  $K$ .

We can make an analogy between rules in a metalanguage to procedures by which natural language is learned. It is uncontentious that the lexicon of a natural language is learned in an a posteriori manner. A child might learn the meaning of "three" through the repeated use of the word in contexts where a group of three apples, balls, fingers etc. are presented. An adult might learn the meaning of "three" using a dictionary. These are just plausible examples, not psycho-linguistic theses. The point is that experience is used when learning a language. Any plausible account of the a priori should take this into account. "Three is the number of apples in the bowl" is an a posteriori sentence by which the meaning of three might be learned, yet this does not affect the apriority of arithmetic statements about the number three. Analogously,  $K$  being a posteriori does not entail that

‘ $N\phi \leftrightarrow \sim \phi$ ’ is a posteriori.<sup>13,14</sup>

At this point I do not wish to make the further positive claim that ‘ $N\phi \leftrightarrow \sim \phi$ ’ is a priori. This would require saying much more about the a priori than I have so far. But as I hope I have shown, the modal and epistemic arguments do not stand.

## 5.5 Concluding Remarks

Peacocke and McCarthy propose criteria for logical terms that involve epistemic notions such as the *a priori* and *epistemically possible situations*. In the case of Peacocke’s criterion, we have seen that in order to maintain the logicity of some standardly accepted logical terms, various seemingly ad hoc modifications need to be made to the basic criterion proposed—to the point where it becomes unclear whether Peacocke provides us with a reliable criterion.

McCarthy proposes a combination of the isomorphism invariance criterion with epistemic notions. I have not discussed the details of his proposal, but rather concentrated on

---

<sup>13</sup>Another option is to make the analogy between rules in the metalanguage and procedures of reference fixing along the lines of Kripke’s theory of reference. In those procedures, the way reference is fixed might well be contingent. But this does not entail that sentences with terms whose reference is fixed contingently are themselves contingent.

<sup>14</sup>McGee uses similar examples to show the insufficiency of invariance. He defines *unicorn negation* as:  $\mathcal{U}\phi =_{Def} (\neg\phi \wedge \text{there are no unicorns})$ , and claims that  $\mathcal{U}$  is intuitively nonlogical, though the operation described by  $\mathcal{U}$  is invariant (McGee, 1996, p. 569f). Since McGee does not ascend to the metalanguage to define  $\mathcal{U}$ , the claim that the operation described by  $\mathcal{U}$  is invariant is problematic. The status of McGee’s definition is not clear to me, but if it is meant as an abbreviation in the object-language, or as a mutual constraint on  $\mathcal{U}$  and *unicorn*, then  $\mathcal{U}$  would be logical only if *unicorn* is, which we have no reason to suppose (see f.n. 12). McGee, however, concludes that not every *connective* that describes a logical operation is itself logical. He thus proposes as a conjecture the following criterion for logical connectives: “A connective is a logical connective if and only if it follows from the meaning of the connective that it is invariant under arbitrary bijections.” (McGee, 1996, p. 578). But, as McGee himself admits, “the notion of something following from the meaning of a word is so terribly slippery.” (*ibid.*) See Chapter 4 for more details on McGee’s criterion.

## CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS

his objection to the original, unmodified criterion of isomorphism invariance. McCarthy shows that invariance criteria allow for definitions of logical terms that depend on contingent or a posteriori assumptions. I have claimed that McCarthy's objection is unsound, as it involves unwarranted assumptions regarding the role of the metalanguage in the definition of logical terms.

I conclude that the epistemic approaches to logicity that were surveyed are untenable. Yet, there is still room for an epistemic component in the concepts of logical terms and logical consequence. As suggested in the beginning of the chapter, there might be an epistemic motivation for the conditions of *Necessity* and *Formality*, as well as for criteria for logical terms that do not explicitly contains epistemic notions.

*CHAPTER 5. EPISTEMIC CRITERIA FOR LOGICAL TERMS*

## Part II

# From Logical Terms to Semantic Constraints



## Chapter 6

# Formality in Logic: From Logical Terms to Semantic Constraints

### 6.1 Introduction<sup>1</sup>

Formality is considered by many to be a central, even definitional feature of logic. Following Tarski, many understand formality to imply that the logical validity of an argument turns on its form. A large part of contemporary discussions on formality is dedicated to the notion of *logical term*, as logical terms are assumed to determine the form of sentences and arguments.

In this chapter I challenge the centrality of logical terms assumed in the contemporary conception of logical consequence. I focus on the model-theoretic tradition, and refer to both principled and relativized accounts of logical terms. In the main part of the chapter (Section 6.3), I propose an alternative framework for logic where logical terms no longer play a distinctive role. The new account employs a new notion of *semantic constraints*. The basic ideas and the philosophical motivation for the new account are presented in this

---

<sup>1</sup>The ideas presented in this chapter are contained in (Sagi, forthcominga).



chapter. The technical framework, including definitions of the main concepts relating to semantic constraints and some results, is presented in the next chapter.

## 6.2 Logical Terms

The thesis of the centrality of logical terms follows from two underlying assumptions, which I will refer to as *the two tenets of formality*:

TF1. The logical validity of an argument is determined by its form.

TF2. The form of an argument is determined by the logical vocabulary and the arrangement of all terms in the argument.<sup>2</sup>

I concentrate on arguments at this point, but of course TF2 can be formulated with respect to forms of sentences as well. “Determined” may be read as *is a function of*. By “arrangement” I mean the pattern of repetitions of terms and their separation by auxiliary devices such as parentheses.<sup>3</sup>

From TF1 and TF2 follows this formulation of the thesis of the centrality of logical terms:

(\*) The logical validity of an argument is determined by the logical vocabulary and the arrangement of all terms in the argument.

---

<sup>2</sup>Compare (Dutilh Novaes, 2012), where eight “tenets of formality” are listed.

<sup>3</sup>Dutilh Novaes indicates that the arrangement of the logical terms in the sentence affects logical validity (Dutilh Novaes, 2012, n. 14). But note that the arrangement of *all* terms in an argument affects logical validity, nonlogical terms as well as logical. Take the two following sentences in first-order logic, which differ only in the arrangement of their nonlogical terms:

$$(1) \forall x(Rxa \rightarrow Rxa)$$

$$(2) \forall x(Rxa \rightarrow Rxb)$$

(1) is a logical truth and (2) isn’t, and this example can be easily used to show that logical *validity* of arguments does not depend solely on the logical vocabulary.

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

There may be various different conditions a valid argument should satisfy (first and foremost: truth-preservation). Whatever these conditions are, (\*) states that whether an argument satisfies them is determined by the logical vocabulary (and the arrangement of all terms).

TF1 is one of the ways by which logical consequence can be said to be formal, as in (Tarski, 1936). Tarski connects TF1 to an epistemic motivation: formality allows the independence of logical consequence from empirical knowledge or knowledge of the objects that are referred to in an argument. I will follow the Tarskian tradition by assuming TF1 for the rest of the chapter (and in this work in general), and specifically that the notion of formality in logic pertains to *forms of arguments (and sentences)*. My account will sit well with the epistemic motivation for TF1, but elaborating on that issue would take us too far afield.

TF1 assumes the notion of a form of an argument. In what follows I challenge the conventional understanding of this notion, which is expressed by TF2, and relates form to logical terms. TF2 is generally assumed rather than argued for. I will give up this assumption in favor of a better motivated, weaker one.

In Chapter 2 I presented two theses about logical terms:

PD. There is a principled distinction between logical and nonlogical terms.

TR. Logical validity is relative to a choice of logical terms, and there is no principled distinction between logical and nonlogical terms.

I also distinguished between a *deep* and a *shallow* sense of *logical terms*. Logical terms in the deep sense are those that are supposedly logical by their “nature”, satisfy some criterion. As noted previously, this understanding of logical terms presupposes PD. Logical terms in the shallow sense are system relative: they are the terms that are completely fixed in a given system.

TF2 (and so its conclusion (\*)) can be understood in two ways, according to whether logical terms are understood in the shallow or in the deep sense. Were one to reject PD

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

and endorse Tarskian Relativism (TR), it would still be possible to maintain TF2 and (\*) when logical terms are understood in the shallow sense. The notions of *form* and *logical consequence* would then be relative as well.

One can reject the view of logic as formal (Read, 1994) or reject TF1 as a correct articulation of that view (see (MacFarlane, 2000)). But one can accept TF1 and still question the thesis of the centrality of logical terms—a possibility that has been neglected in contemporary literature. The thesis that has mainly been up for debate is PD, that there is a principled distinction between logical and nonlogical terms. TR (and its relative, the pragmatic approach) is a neat way of rejecting PD while holding on to the two tenets of formality—but it still assumes the centrality of logical terms. I propose to consider alternatives to the thesis of the centrality of logical terms, without giving up on formality or on TF1. This means we should let go of TF2 and that the form of an argument (or a sentence) requires a strict division between logical and nonlogical terms, without altogether giving up on the notion of a form.

Thus, at this point I propose to take a step back, and see what can be achieved without assuming a division between logical and nonlogical terms. We can start by looking into a possible motivation for the logical/nonlogical dichotomy. We noted earlier that logical terms are those terms that *are* held fixed in the system (this is the shallow sense of “logical terms”) or those that *should* be held fixed (understood in the deep sense). Analogously, nonlogical terms *are* variable in the system, or are those terms that *should* be variable. Whichever reasons we may have for fixing some terms and not others might possibly also justify fixing some terms to *some extent*, keeping them also variable to some extent. Let’s take an example. One might claim that logical consequence has “epistemic virtues” such as certainty or aprioricity and would require that only terms that yield an “epistemically virtuous” consequence relation should be fixed under all interpretations. These might be terms whose extensions we know with complete certainty. But in the same vein, one can fix other terms *partially*, to an extent that keeps logic “epistemically virtuous”. Or, in other

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

words, we can limit the variability of some terms without fixing them completely.

This line of thought leads to the following, if somewhat vague, alternative to TF2, at least when understood in a “shallow” way:

TF2'. The form of an argument is determined by what is held fixed in the argument under all interpretations.

We note that the syntactic categorization of terms (into predicates, connectives etc.) and their overall arrangement in the argument are some of the things that are held fixed. It is clear from the discussion above that TF2' is in line with the epistemic motivation for TF1 implied by Tarski.

If TF2' is to give us a better understanding of the notion of a form of an argument, we need to first define the class of (admissible) interpretations. Usually, this is done by specifying the logical terms. TF2', which is no longer centered on logical terms, may now inspire us to search for new ways of fixing things in an argument or a sentence. We can go beyond the limiting way of looking at form through the logical/nonlogical distinction between terms and adopt a more general view instead. In what follows, I describe a formal framework and provide examples of ways terms can be fixed in different manners and to various extents. We will then be able to restate the idea behind TF2' in a clearer and more precise manner.

### 6.3 Semantic Constraints

Acknowledging that within the standard conception of logic what matters to the form of an argument is what is considered to be fixed opens the subject to a range of new possibilities. In what follows I explore one of the options in that range. I devise a notion of form that does not rely on a distinction between logical and nonlogical terms: the terms of the language will not be strictly divided to those that have to do with its form and those that haven't.

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

Instead of trying to distinguish between fixed and non-fixed terms, I consider fixing terms partly to various extents, or more generally, fixing or constraining parts of the language in different ways.

Fixing a term as logical may be viewed as restricting the admissible models for the language.<sup>4</sup> Now take, for instance, the terms *allRed* and *allGreen*.<sup>5</sup> These are paradigmatic cases of nonlogical terms in mainstream logic. There are good reasons for not fixing the extensions of color terms completely. But we could fix their mutual dependencies, and have a rule in our system that says that their intersection is empty in all models. A rule like this, I contend, is not essentially different from a rule fixing the interpretation of a logical term. In both cases, there is a rule that consists in restricting admissible models. We may say that whereas the interpretations for logical terms are completely fixed, other terms may be *partly* fixed.

Let us refer to all such rules on the admissible models for a language as *semantic constraints*. Logical terms (or more precisely, rules defining logical terms) are merely a special case of semantic constraints, while *all* the semantic constraints in a system are involved in determining logical consequence.

I will assume that a language L consists of a set of terms (the atomic expressions in L) and a set of phrases (the set of all meaningful expressions in L). A phrase consists of a string of terms and perhaps auxiliary devices such as parentheses. Phrases are interpreted by the models for the language.

---

<sup>4</sup>See e.g. (Varzi, 2002). Varzi contends that *all* terms receive their meaning through the restriction of admissible models, and there is therefore no principled distinction between logical and nonlogical terms: “What are the reasons for maintaining that the distinction between the logical and the extra-logical is up for grabs? Broadly speaking, my reasons stem from the consideration that all bits of language get their meaning fixed in the same way, namely, by choosing some class of models as the only admissible ones.” (Varzi, 2002, p. 199).

<sup>5</sup>Consider *allRed* and *allGreen* as primitive terms in the object language which stand for *red all over* and *green all over*.

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

In a model-theoretic framework, semantic constraints for a language  $L$  will be sentences in the metalanguage, usually with a universal quantification ranging over all models (domains and interpretation functions) which I will omit. For technical purposes I require that each semantic constraint refers to, or limits the interpretations of, finitely many terms in  $L$ .<sup>6</sup> We also make use of the constants  $T$  and  $F$  for the truth values. Let a model  $M$  be a pair  $\langle D, I \rangle$  where  $D$ , the domain, is a non-empty set, and  $I$  an interpretation function for  $L$ . In this general, nonstandard setting, what we mean by an interpretation function is a function that takes as arguments phrases in  $L$  and whose values are objects in the set-theoretic hierarchy with  $D \cup \{T, F\}$  as ur-elements.

The previous paragraph characterizes semantic constraints in a *shallow* sense, as constituents in a formal system. My aim in what follows is to lay out the framework of semantic constraints, not give a demarcation of the “correct” semantic constraints. Thus, I will not offer solutions to the questions of logicity that have been broached in the term-based approach. But moving to a more general framework may dissolve some of the previous questions, and change the considerations in others. The principal benefit is the removal of an unfounded assumption. The proposed framework still allows for the more restricted, term-based, systems; but in this setting, the exclusivity of such systems in defining logical consequence is not taken for granted.

Moving on to some examples, the semantic constraint for standard conjunction of sentences will be:

$$(\wedge): I(\varphi \wedge \psi) = T \Leftrightarrow I(\varphi) = T \text{ and } I(\psi) = T$$

where  $\wedge$  is in  $L$ , and  $\varphi$  and  $\psi$  range over sentences in  $L$ .<sup>7</sup> From now on I will assume that a term to which an interpretation function is applied in a semantic constraint is in  $L$ . A way

---

<sup>6</sup>This assumption is needed for Proposition ?? in the following chapter.

<sup>7</sup>The semantic constraint that includes conjunction of open formulas as in standard first-order logic is somewhat more elaborate and is not presented in this chapter for ease of exposition. See Chapter 7.2.2 for the full set of constraints for standard first-order logic.

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

of reading the above definition is: a model  $\langle D, I \rangle$  is admissible in the semantics only if it satisfies  $(\wedge)$ . We may also wish to restrict the class of models with respect to terms in the language without fixing their extension absolutely, for instance, by:

- $0 \in I(\text{naturalNumber})$ ,

restricting the term *naturalNumber*, so that 0 is always in its extension.<sup>8</sup> In addition, semantic constraints can restrict classes of models by relating the interpretation of some terms to others:

- $I(\text{even}) \cap I(\text{odd}) = \emptyset$ ,

restricting *even* and *odd*.

We can also treat the division of terms into syntactic categories as semantic constraints, so that the constraint

- $I(\text{big}) \subseteq D$

tells us that *big* should be interpreted as a unary predicate. This may seem more like a syntactic rather than a semantic constraint, but there is no need for hard and fast distinction between syntax and semantics here. The above item is a constraint on interpretations of the language, and is thus considered here as a semantic constraint.

The semantic constraint mentioned informally previously can be formulated thus:

- $I(\text{allRed}) \cap I(\text{allGreen}) = \emptyset$ .

We can likewise have such a constraint for any pair of color terms.

More examples for semantic constraints include:

---

<sup>8</sup>This semantic constraint also happens to restrict models by imposing the requirement that 0 will be a member of every domain.

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

- $I(\text{bachelor}) \subseteq I(\text{unmarried})$
- $I(\text{wasBought}) = I(\text{wasSold})$
- $I(\text{John}) \in D$

A recursive definition of truth or satisfaction can be incorporated in a system of semantic constraints. From now on I will assume that  $L$  is a first-order language, and that the systems of constraints we are dealing with include the usual first-order semantics, augmented with further constraints as those above.<sup>9</sup>

The consequence relation for  $L$  is determined by a collection of semantic constraints. Let  $\Delta$  be a set of constraints including those mentioned above.  $\Delta$  fixes a class of models for  $L$ : those satisfying all the semantic constraints in  $\Delta$ . An argument  $\langle \Gamma, \varphi \rangle$  (where  $\Gamma$  is a set of sentences in  $L$  and  $\varphi$  is a sentence in  $L$ ) is *valid* in  $\Delta$  if for any  $\Delta$ -model (an admissible model for  $\Delta$ , satisfying those constraints), if all the sentences in  $\Gamma$  are true, then  $\varphi$  is true as well. Let  $\models_{\Delta}$  be the associated consequence relation, so, for instance we have:  $\text{bachelor}(\text{John}) \models_{\Delta} \text{unmarried}(\text{John})$ .

On this view, the collection of semantic constraints represents, in a very general sense, the form, or structure, of the language. We might not commit to giving the exact extension of *allRed* in every model, and yet commit to its extension having no common member with the extension of *allGreen* in all models. The narrow logical/nonlogical view of terms doesn't allow for such a mutual half-fixing of terms. But just as we constrain *allRed* to be

---

<sup>9</sup>By  $L$  being a first-order language I mean that the language consists of individual constants, predicates of each finite arity, individual variables, first-order quantifiers, sentential connectives and auxiliary symbols (parentheses and comma). I also assume that well-formed formulas can be defined recursively. In present terms, only closed expressions (closed formulas and singular expressions) are properly considered as phrases. Note that in a recursive definition of well-formed formulas there is no need to fix the logical terms of the language. For example, there can be a clause saying that if  $\varphi$  and  $\psi$  are well-formed formulas and  $C$  is a binary sentential connective, then  $\varphi C \psi$  is a well-formed formula, regardless of whether  $C$  is fixed as a logical term.



## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

a predicate, we may also constrain it to be disjoint from *allGreen*. Even if a term is not to be completely fixed and considered as a logical term in a system in the standard sense, its range of interpretations can still be limited, and this option should have place in a theory of logical reasoning.

We receive a view where *form* is treated far more generally than it usually is, and yet the generalization is a very natural one.<sup>10</sup> Logical terms still have a special place in our system, as they lie at the end of a spectrum—they are the terms whose denotations are completely fixed (i.e. their interpretation is a function of the domain of the model). But formality is no longer subsumed under the issue of logical terms. The notion of form I am proposing has to do with constraints on the language and on the interpretation of arguments rather than with specific terms.

So our final reformulation of TF2 will be:

(TF2\*) The form of an argument in a language L is determined by a set of semantic constraints for L and the arrangement of terms in the argument.

Together with TF1, we obtain the following conclusion:

(\*)\* The logical validity of an argument in a language L is determined by a set of semantic constraints for L and the arrangement of terms in the argument.

---

<sup>10</sup>Views of formality that are similar in their generality can be found in the literature. We might not go so far as Chomsky and claim that “To say that a relation is formal is to say nothing more than that it holds between linguistic expressions” (Chomsky, 1955, p. 39). We might require, in the spirit of Carnap, that a formal relation must be determined by explicitly given rules of the language (see (Carnap, 1937, p. 186)), e.g. in a system of constraints.

## 6.4 Meaning postulates, Cross-Term Restrictions and Semantic Constraints

Semantic constraints resemble ideas already existing in the literature. Let us take a brief look at the concept of *meaning postulates* in the work of Carnap and later in Montague. In a formal system, meaning postulates are sentences in the object language, fixing the relation between terms. That is, models (*state-descriptions* in the case of Carnap) considered for logical consequence (*L-implication* or *analyticity* for Carnap) are only those that satisfy the meaning postulates (Carnap, 1947, p. 226), (Montague, 1974b, pp. 263-264). Let us observe an example from Montague. The meaning postulate

$$\Box[seek'(x, P) \leftrightarrow try - to'(x, \hat{find}'(P))]$$

fixes the relation between the expressions *seek* and *try-to find*, as can be observed without going into the details of Montagovian notation.

Notably, the semantic constraints suggested above look like a variation of meaning postulates. However, for both Carnap and Montague meaning postulates and logical terms reside in two separate parts of the system. Carnap introduces meaning postulates for his explication of *analyticity*, which he distinguishes from a narrower concept of *logical truth* that does not depend on meaning postulates. For his notion of *logical truth*, Carnap relies on a distinction between logical and nonlogical (descriptive) terms.<sup>11</sup> Montague, too, relies on a distinction between logical and nonlogical terms. But by contrast to Carnap, he lets meaning postulates affect logical consequence as well. Yet, he does not consider the rules for logical terms among the meaning postulates, although both alike determine the extension of logical consequence. By contrast to both Carnap and Montague, in the framework

---

<sup>11</sup>In (Carnap, 1947), logical terms are given in lists, and the distinction is considered to be a matter of convention. Previously, in (Carnap, 1937), logical terms were given in a syntactic fashion as such that sentences constructed by them are “determinate”. Both cases are of a system-relative notion of logical terms (see *ibid*, p. 178f and the discussion in Chapter 2).

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

presented here there is no fundamental distinction between restrictions for logical terms and restrictions for other terms. Consequently, Carnap's distinction between *analyticity* and *logical validity* is discarded. The question of whether there is place for an interesting notion of analyticity, different from Carnap's, that can be contrasted to logical validity on this approach, will have to be left out of the present discussion.

So we can sum up two important differences between semantic constraints and meaning postulates as they appear in Carnap and Montague. First, semantic constraints are formulated in the metalanguage and meaning postulates in the object language. This means that semantic constraints will have, in general, greater expressive power, while meaning postulates can be considered as axioms or as a theory for the language. Secondly, and relatedly, by contrast to Carnap and Montague's systems, there is no fundamental difference between the rules or definitions of logical terms and other semantic constraints. Semantic constraints mark the form of a sentence or an argument by holding parts of it fixed, and therefore are justifiably a factor in a language's logical consequence relation. They are a wide category under which logical terms are also subsumed.<sup>12</sup>

One of the main criticisms of meaning postulates or similar "semantical rules" is their *ad hoc* nature and the lack of explanatory power. Thus, Quine claims that "Semantical rules are distinguishable, apparently, only by the fact of appearing on a page under the heading 'Semantical Rules'; and this heading is itself then meaningless" (Quine, 1951, p. 33). Etchemendy claims that meaning postulates are chosen only on the account of what seems logical true, making the definition of logical truth circular (Etchemendy, 1990, p. 73).

---

<sup>12</sup>I have only referred to meaning postulates *vis-à-vis* their role in the formal system, and disregarded their role in giving a semantic theory for natural language, where they are typically used by linguists. Semantic constraints can be used in formal systems for various purposes, among which is the investigation of natural language, but I am not arguing here for the efficacy of using semantic constraints in linguistics. The point here is that form need not be determined through a strict dichotomy between two types of terms, the logical and the nonlogical.

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

The criticisms regarding meaning postulates seem to apply to semantic constraints just as well. Nonetheless, these criticisms are besides the point in the present discussion. The reason is that I have made no claims about the way semantic constraints are chosen, which need not be arbitrary. My point, in the time being, is that instead of either completely fixing a term or not fixing it at all, there is room to consider partly fixing some terms. I return to Quine’s criticism of meaning postulates in Chapter 10.2.

Semantic constraints also have much in common with *cross-term restrictions* discussed in (Etchemendy, 1990, Chapter 5). In the framework with which Etchemendy deals, the language is interpreted by *d-sequences*: they are interpretation functions similar to the ordinary ones we use in models, but the objects or sets of objects they assign are all taken from one universe—a predetermined satisfaction domain. Cross-term restrictions are constructed as a means for creating a semantics with varying domains. I shall not go into the details of Etchemendy’s use of cross-term restrictions, but ultimately, what cross-term restrictions do is constrain the interpretation of terms in the language by excluding some *d-sequences* from the semantics. In this they are similar to meaning postulates and semantic constraints. In the end, however, Etchemendy rejects cross-term restrictions. He claims that they conflict with two important principles of logical consequence, of substitution and persistence, to be explained below. This conflict might be used as criticism against semantic constraints as well. Here I shall present the criticism and my response in brief, and my full solution is presented in Chapter 7.3.4.

The *principle of substitution*, taken from Bolzano, says that a logical truth remains a logical truth after uniform substitution of the non-fixed terms according to their grammatical categories (the same goes for logically valid arguments). The *principle of persistence* says that the property of being logically true with respect to a given set of fixed terms should persist through simple expansions of the language (Etchemendy, 1990, p. 31). Similarly, the principles of substitution and persistence can be formulated with respect to the logical validity of arguments. So, adding a term to a language should not change the logical

status of sentences or arguments constructed by terms in the original language. Now, as Tarski famously pointed out, if the principle of substitution is taken as a sufficient condition for logical consequence (or logical truth), persistence is violated: logical consequence becomes dependent on the richness of the language (Tarski, 1936). Tarski’s definition of logical consequence using satisfaction is constructed to solve this problem.

Etchemendy claims that when using cross-term restrictions, substitution and persistence cannot both be held (Etchemendy, 1990, p. 70). I shall not go into the details of Etchemendy’s examples. But, *prima facie*, the same problem should arise for semantic constraints. Very simply: accepting just the constraint  $I(allRed) \cap I(allGreen) = \emptyset$  and the standard constraints of predicate logic makes the sentence  $\neg\exists x(allRed(x) \wedge allGreen(x))$  logically true, while its substitution instance  $\neg\exists x(Giant(x) \wedge allGreen(x))$  is not logically true. Thus, it would seem, we violate substitution, unless we exclude some predicates from the language, in which case we violate persistence. The problem of substitution is dealt with and solved in Chapter 7.3.4. The basic idea is that the relevant categories of substitution are no longer the traditional grammatical categories, but rather semantic categories containing only terms that behave the same in the system. *allRed* and *Giant* do not belong to the same category (as defined on p. 141), and thus should not be considered intersubstitutable.

## 6.5 Conclusion

Contemporary debates in logic rely on the assumption that arguments have forms, and that an argument’s form is determined by its logical vocabulary: the so called “tenets of formality”. In this chapter I proposed a view of logic that relies on forms, but takes a step towards breaking free of the conception of form as constituted by logical terms.

I contend that a false dichotomy results when taking logical terms to be those that have to do with form, and nonlogical terms to be those that do not. Form has to do with everything that is held fixed in a sentence or an argument. From this standpoint, new logical

## CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC CONSTRAINTS

frameworks become available. Any *semantic constraint* on a language, a way of explicitly fixing terms of the language in some manner or to some degree, contributes to the form of sentences in the language. We may hope that the general view of formality that emerges will open the field of logic to a variety of systems that have not hitherto been considered.

Is there one correct system of semantic constraints for logical consequence? The framework I propose has the permissive spirit of relativism, but one can accept the theory of semantic constraints and deny relativism. It can be claimed that there is one set of constraints which determines logical consequence. Alternatively, one can be skeptical of the search for the “correct” set of constraints as much as one can be of the “correct” set of logical terms. This question is left open at this stage, but we should note that even in a relativistic framework there could be theoretical preferences for formal systems that fix certain terms to some degree and not further.<sup>13</sup> In Chapter 8 I suggest considerations for choosing semantic constraints and “measuring” their logicity.

---

<sup>13</sup>For instance, we may fix the dependencies between color terms but refrain from fixing their extensions completely. There may be principled reasons for this (along the lines of the epistemic motivation mentioned earlier), but also various pragmatic reasons. We might not want to commit to a fixed extension and would like to consider different possibilities, we might want to keep things simple, etc.

*CHAPTER 6. FORMALITY IN LOGIC: FROM LOGICAL TERMS TO SEMANTIC  
CONSTRAINTS*

## Chapter 7

# Semantic Constraints: The Framework

This chapter sets up the basic framework of semantic constraints. First, the three components of the framework are presented: the language, the models for the language, and the system of semantic constraints for the models of the language. Next, standard propositional and first order logic are put into the framework. Nonstandard examples are then provided, demonstrating how the framework extends the conventional understanding of a logic. Then, in the main part of the chapter, various definitions and propositions regarding the framework are presented and discussed. These pertain to the nature of semantic constraints, to properties and relations terms may hold in such a system, to the division of terms into categories and families, and to the schematization of arguments in the framework.



## 7.1 Language, Models, and Semantic Constraints

### 7.1.1 The language L

A language L consists of *terms* and *phrases*. Terms are primitive, and phrases are strings of terms, possibly including auxiliary devices (such as parentheses).

### 7.1.2 Models

At this stage we deal with basic model-theoretic semantics for extensional languages. A model for L,  $M = \langle D, I \rangle$ , consists of a non-empty set  $D$  (the *domain*) and an interpretation function  $I$ .  $I$  assigns values to the phrases in L (not all strings of terms are phrases!). The values are received set-theoretically from  $D$  and the set of truth values. We deal here with a set of two truth-values,  $\{T, F\}$ . So for each phrase  $p$  in the domain of the function  $I$ ,  $I(p)$  is an element in the set-theoretic hierarchy with  $D \cup \{T, F\}$  as ur-elements.

### 7.1.3 Semantic constraints

A *semantic constraint* for L is a sentence in the metalanguage that somehow constrains or limits the admissible models for L. Semantic constraints include implicit universal quantification over models (domains and interpretation functions). At this point I leave the notion of a semantic constraint as general as possible. I will add further assumptions regarding semantic constraints where they are needed.

Let  $\Delta$  be a set of semantic constraints. A  $\Delta$ -model is an admissible model by  $\Delta$ , i.e. a model abiding by the constraints in  $\Delta$ .

As we deal here with models set up on a classical set-theoretic foundation, the metalanguage used in all the examples includes, but is not confined to, the language of set theory.

### 7.1.4 The status of the domain

For various reasons, we might like to constrain the domains of models, and not just the way terms are interpreted. For uniformity, we can add a term which denoted the domain in each model, so that constraints on the domain are tantamount to constraints on that term. Indeed, in standard first order languages the term  $\forall$  (as exhibited below), denotes the singleton containing the domain, so that every constraint on  $D$  is a constraint on  $I(\forall)$  and vice versa.<sup>1</sup> For convenience, we shall add to each language a (quasi) term  $\mathcal{D}$  for the domain. The significance of this treatment is that the definitions and propositions about terms below will then apply to the domain as well. For this purpose, we postulate that for any language  $L$  and every model  $\langle D, I \rangle$  for  $L$ ,  $I(\mathcal{D}) = D$ . The domain will of course keep its status as providing the building blocks of the interpretations of all other terms.

## 7.2 Examples of Constraints

### 7.2.1 Propositional logic

Let us begin with examples of basic constraints terms from standard propositional logic. The terms of  $L$  are:  $\{p_i\}_{i \in \mathbb{N}} \cup \{\neg, \wedge, \vee, \rightarrow\}$ . Each of the following constraints deals with one phrase in the language. A constraint that pertains to an arbitrary phrase of some sort (e.g. the first one below) should be regarded as a schema for infinitely many constraints, each referring to one term. We assume the usual definition for *well-formed formula (wff)* for propositional logic to facilitate the schematic formulation of constraints. (These are *definitions*, not constraints, and are not listed below).

1.  $I(p_i) = T$  or  $I(p_i) = F$  for all  $i \in \mathbb{N}$ .

---

<sup>1</sup>In some presentations of first-order logic, the universal quantifier is presented as a *parameter*, not a *constant*: signifying the variability of the domain, see (Enderton, 1972, p. 68).

2.  $I(\neg) = f_{\neg}$  where  $f_{\neg}$  is the negation function from truth-values to truth-values:  $f_{\neg} = \{(v_1, v_2) : v_1, v_2 \text{ are truth-values, } v_2 = T \text{ iff } v_1 = F\}$ .
3.  $I(\wedge) = f_{\wedge}$  where  $f_{\wedge}$  is the conjunction function from pairs of truth values to truth-values:  $f_{\wedge} = \{((v_1, v_2), v_3) : v_1, v_2, v_3 \text{ are truth-values, } v_3 = T \text{ iff } v_1 = v_2 = T\}$ .
4.  $I(\vee) = f_{\vee}$  where  $f_{\vee}$  is the disjunction function from pairs of truth values to truth-values:  $f_{\vee} = \{((v_1, v_2), v_3) : v_1, v_2, v_3 \text{ are truth-values, } v_3 = F \text{ iff } v_1 = v_2 = F\}$ .
5.  $I(\rightarrow) = f_{\rightarrow}$  where  $f_{\rightarrow}$  is the material implication function from pairs of sentences to truth-values:  $f_{\rightarrow} = \{((v_1, v_2), v_3) : v_1, v_2, v_3 \text{ are truth-values, } v_3 = F \text{ iff } v_1 = T \text{ and } v_2 = F\}$ .
6. For any wff  $s$ ,  $I(\neg s) = I(\neg)(I(s))$ .
7. For  $c = \wedge, \vee, \rightarrow$  and any two wffs  $s_1, s_2$ ,  $I(s_1 c s_2) = I(c)((I(s_1), I(s_2)))$ .

### 7.2.2 Standard first order predicate logic

We can extend the logic to standard first order predicate logic in a natural way. The terms are as above, to which we add individual constants, relation and function symbols, variables and quantifiers.

To deal with variables, we use *variable assignments* in the usual way. A variable assignment  $\sigma$  for a model  $M$  is a function from variables to  $D$ . We start with the valuations of  $I$  on individual constants and function and relation symbols. Then we use interpretation functions and variable assignment pairs (the pair  $(I, \sigma)$  will be signified by  $I_{\sigma}$ ) to valuate open phrases. We then define  $I$  for closed phrases.

We have characterized semantic constraints as limiting the admissible models for a language. Models are defined by  $D$  and  $I$ , and  $I_{\sigma}$  is not part of the model *per se*. However,  $I_{\sigma}$  needs to be constrained in a sense – the way it is constrained will affect the way  $I$  is

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

constrained. The items concerning  $I_\sigma$  can be seen as limiting the admissible models for the language indirectly, given that there are also constraints linking  $I_\sigma$  to  $I$ . But it will be more appropriate to consider the items pertaining to  $I_\sigma$  as a *definition* of  $I_\sigma$  as it appears in the constraints involving  $I$ , and as auxiliary to them. If the items involving  $I_\sigma$  were considered to be semantic constraints on their own, we would have the problematic situation of some constraints making sense only when others are in place.<sup>2</sup>

We use here the conventional recursive definitions of a *well-formed formula (wff)* and of *singular term* in predicate logic. Again, each of the following constraints deals with one phrase in  $L$ , so most of the following items should be regarded as schemas for infinitely many constraints.

8.  $I(c) \in D$  (for an individual constant  $c$ ) (i.e. individual constants are singular phrases).
9.  $I(f)$  is a function from  $D^n$  to  $D$  (for an  $n$ -place function symbol  $f$ ,  $n \geq 1$ ).
10.  $I(R) \subseteq D^n$  for an  $n$ -place relation symbol  $R$ ,  $n \geq 1$ .
11.  $I(\forall) = \{D\}$
12.  $I(\exists) = \{A \subseteq D : A \neq \emptyset\}$
13.  $I(\varphi) = v$  where  $\varphi$  is a wff with no free variables and  $v$  a truth value such that for every variable assignment  $\sigma$ ,  $I_\sigma(\varphi) = v$ .

Now the following items are not properly constraints, but should be considered as auxiliary definitions for constraint 13.

- $I_\sigma(\alpha) = I(\alpha)$  where  $\alpha$  is an individual constant, function symbol, relation symbol or a sentential connective.

---

<sup>2</sup>This would make the semantic constraints contextual (i.e. affected by context). See f.n. 20 for problems that arise from contextuality.

- $I_\sigma(x) = \sigma(x)$  where  $x$  is a variable.
- $I_\sigma(f(t_1, \dots, t_n)) = I_\sigma(f)((I(t_1), \dots, I(t_n)))$  where  $f$  is a function symbol and  $t_1, \dots, t_n$  are singular terms.
- $I_\sigma(R(t_1, \dots, t_n)) = v$  where  $R$  is a relation symbol,  $t_1, \dots, t_n$  are singular terms and  $v$  is a truth-value such that  $v = T$  iff  $(I(t_1), \dots, I(t_n)) \in I(R)$ .
- $I_\sigma(\neg\varphi) = I(\neg)I_\sigma(\varphi)$  where  $\varphi$  is a wff.
- $I_\sigma(\varphi \text{ } c \text{ } \psi) = I(c)(I_\sigma(\varphi), I_\sigma(\psi))$  where  $c = \wedge, \vee, \rightarrow$  and  $\varphi$  and  $\psi$  are wffs.
- $I_\sigma(\varphi^{x_1, \dots, x_n}) = \{(d_1, \dots, d_n) \in D^n : I_{\sigma[x_i/d_i]_{1 \leq i \leq n}}(\varphi) = T\}$  where  $\varphi$  is a wff.
- $I_\sigma(Qx\varphi) = v$  where  $Q$  is a quantifier,  $\varphi$  is a wff and  $v$  is a truth-value such that  $v = T$  iff  $I_\sigma(\varphi^x) \in I(Q)$

### 7.2.3 Constrained FOL

Alternative systems of constraints can add constraints to those of propositional or first order logic, and can also subtract some of the standard constraints. For instance, in an alternative system, there may be terms for which there is no constraint on their grammatical category: in one model they are evaluated as individual constants and in another as predicates. But having been put together in certain phrases, the constructed phrases can then be constrained, e.g. to be sentences. Such cases can be conceived of as portraying partial information of a given language.

Of course, the framework of semantic constraints is not committed to having standard logic as its basis. Intuitionistic, relevant and other non-classical logics can be accommodated. The framework is general enough to account for any possible language that has terms, phrases, and an appropriate system of models as defined above. Extensions to intensional systems are also possible, see Chapter 9.

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

Below are some examples of constraints meant to be added to the system of first order logic. All the terms mentioned are primitive, and are unary predicates in the language.<sup>3</sup>

$$14. I(bachelor) \subseteq I(unmarried)$$

$$15. I(even) \cap I(odd) = \emptyset$$

$$16. I(prime) = \{n : n \text{ is a prime number}\}$$

$$17. 0 \in I(naturalNumber)$$

$$18. I(a_1) \cap I(a_2) = \emptyset \text{ for any pair of distinct "color terms" } a_1 \text{ and } a_2 \text{ (where the "color terms" are a finite set including } allRed, allGreen, allYellow \text{ etc.)}$$

$$19. I(big) \cap I(small) = \emptyset$$

$$20. I(a) \subseteq I(extended) \text{ where } a \text{ is a color term.}$$

$$21. I(big) \subseteq I(extended)$$

$$22. I(small) \subseteq I(extended)$$

$$23. I(wasBought) = I(wasSold)$$

Most of the constraints just added set a relation between two terms. There is one constraint (16) which completely fixes the interpretation of a given term. That constraint also implies the following constraint:

$$\bullet \{n : n \text{ is a prime number}\} \subseteq D$$

limiting the range of possible domains. The fourth constraint only partially fixes the interpretation of a term. It says that the number 0 is always a member of the interpretation of *naturalNumber*. This constraint also implies a constraint on the domain:

---

<sup>3</sup>This means that constraint 10 already applies to them.

- $0 \in D$

Below are some further examples of possible semantic constraints. We assume that  $john, bachelor, R, a, b, c, d$  and  $Q$  are all terms in  $L$ , not necessarily divided into grammatical categories.

24.  $I(john) \in I(bachelor)$
25.  $I(R)$  is a symmetric binary relation.
26.  $I(abc)$  is a sentence.
27.  $I(d) \neq I(\wedge)$
28.  $I(or) \in \{f_{\vee}, f_{\underline{\vee}}\}$  where  $f_{\vee}$  is as in item (4) above, and  $f_{\underline{\vee}}$  is the xor function from pairs of truth values to truth values:  $f_{\underline{\vee}} = \{((v_1, v_2), v_3) : v_1, v_2, v_3 \text{ are truth values, } v_3 = F \text{ iff } v_1 = v_2 = T \text{ or } v_1 = v_2 = F\}$ .
29.  $I(Q) = \{A \subseteq D : 0 \in A\}$  ( $Q$  is a nonstandard quantifier.)

## 7.3 Basic Notions in the Framework

### 7.3.1 Constraining a phrase and non-trivial constraints

**Definition 5** A semantic constraint  $C$  *constrains a set of phrases  $P$  with respect to the set of constraints  $\Delta$*  if there is a  $\Delta$ -model  $M = \langle D, I \rangle$  such that:

1. There is a  $\Delta \cup \{C\}$ -model  $M' = \langle D, I' \rangle$  such that for any phrase  $r \notin P$ ,  $I'(r) = I(r)$ .
2.  $M$  itself is not a  $\Delta \cup \{C\}$ -model.

A semantic constraint  $C$  *constrains a phrase  $p$  with respect to the set of constraints  $\Delta$*  if it constrains  $\{p\}$  with respect to  $\Delta$ .

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

The idea is that  $C$  constrains a set of phrases if it excludes some model without restricting the interpretations of other phrases.

**Definition 6** A constraint  $C$  is *non-trivial* if there is a  $\emptyset$ -model that is not a  $\{C\}$ -model.

A constraint  $C$  is *non-trivial* with respect to a set of constraints  $\Delta$  if there is a  $\Delta$ -model that is not a  $\Delta \cup \{C\}$ -model.

### 7.3.2 Sentences and singular phrases

In the previous section we made use of the standard definitions of *sentence* and *singular term* in propositional and first order predicate logic. These definitions allow us to distinguish special sets of phrases in the language in a recursive way. From a semantic point of view, what characterizes each of these sets is the type of values their members can receive from the interpretation function. That is, sentences are characterized as having truth values as their possible values, and singular terms as having members of the domain as their possible values. In a general framework of semantic constraints, uncommitted to this or that logic, those characteristics can become defining features.

Thus, we can define a *sentence* with respect to a set of semantic constraints  $\Delta$ :

**Definition 7** (Sentence) Let  $\Delta$  be a set of constraints. A phrase  $s$  in  $L$  is a *sentence* (w.r.t  $\Delta$ ) if for each  $\Delta$ -model  $M = \langle D, I \rangle$  for  $L$ ,  $I(s) = T$  or  $I(s) = F$ .

The definition of a  $\Delta$ -*sentence* is wholly semantic. In a similar manner, we can define *singular phrases* in a way that does not depend on the particular syntactic structure of the language:

**Definition 8** (Singular phrase) Let  $\Delta$  be a set of constraints. A phrase  $t$  in  $L$  is a *singular phrase* (w.r.t  $\Delta$ ) if for each  $\Delta$ -model  $M = \langle D, I \rangle$  for  $L$ ,  $I(t) \in D$ .



### 7.3.3 Relations between terms: determinateness and dependency

There are some relevant notions in the logic literature, such as *definability* and related notions that pertain to the relation between terms, that can be adapted to the framework of semantic constraints. One can define *definability* in either a semantic or a syntactic setting. Let us consider Tarski's work, in which he dealt with different restrictions on terms in a syntactic setting (Tarski, 1934).<sup>4</sup> There, notions of *definability* and *independence* are considered with respect to extra-logical terms. Tarski defines *definability* thus:

Let ' $a$ ' be some extra-logical constant and  $B$  any set of such constants. Every sentence of the form:

$$(I) (x) : x = a. \equiv .\phi(x; b', b'', \dots),$$

where ' $\phi(x; b', b'', \dots)$ ' stands for any sentential function which contains ' $x$ ' as the only real variable, and in which no extra-logical constants other than ' $b'$ ', ' $b''$ ', ... of the set  $B$  occur, will be called [...] a *definition of the term ' $a$ ' by means of the terms of the set  $B$* . We shall say that the term ' $a$ ' is *definable by means of the terms of the set  $B$  on the basis of the set  $X$  of sentences*, if ' $a$ ' and all the terms of  $B$  occur in the sentences of the set  $X$  and if at the same time at least one possible definition of the term ' $a$ ' by means of the terms of  $B$  is derivable from the sentences of  $X$ . ((Tarski, 1934) p. 299)

For our purposes we can define an analogous notion of *determinateness*. By contrast to Tarski, we consider in the definition all phrases of  $L$  (not just the terms).

**Definition 9** (Determinateness) A phrase  $p$  is *determined by the set of phrases  $B$*  (w.r.t.  $\Delta$ ) if for any two  $\Delta$ -models  $M = \langle D, I \rangle$  and  $M' = \langle D', I' \rangle$ , if  $I(b) = I'(b)$  for all  $b \in B$  then

---

<sup>4</sup>The basic framework for Tarski is the system of the *Principia*. The definitions below can easily be transformed to a model-theoretic setting, where "derivability" is changed to "logical consequence" and the proper adjustments are made.

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

$$I(a) = I'(a).^5$$

Note that if  $a$  is definable by the terms in the set  $B$  on the basis of the set  $X$  of sentences, then  $a$  is determined by  $B \cup \{\mathcal{D}\}$  in a system of constraints that includes the constraint  $I(\varphi) = T$  for all  $\varphi \in X$ ,<sup>6</sup> but not necessarily vice versa (definability requires expressibility in the object language that is not required by determinability).

We can now define the *logical terms* (the “completely fixed” terms) of the system. Looking at the semantic constraints for propositional logic and for predicate logic, we can see that the standard logical terms are determined by the domain, and in that sense they are *fixed*. But there is a stronger sense by which they are fixed: any phrase containing them is determined by the domain and the other terms in the phrase. When we assume that the language is compositional, these two senses become equivalent. We use the weaker sense to define logical terms, and define compositionality below.

**Definition 10** (Logical term) A term  $a$  is a *logical term* (w.r.t.  $\Delta$ ) if it is determined (w.r.t.  $\Delta$ ) by the domain, i.e. by  $\{\mathcal{D}\}$ .

**Definition 11** (Compositionality) A language  $L$  is *compositional* (w.r.t.  $\Delta$ ) if each phrase  $p$  consisting of the terms  $a_1, \dots, a_n$  and auxiliary symbols is determined by  $\{\mathcal{D}, a_1, \dots, a_n\}$  (w.r.t.  $\Delta$ ).<sup>7</sup>

---

<sup>5</sup>The notion of *determinateness* by is the same as that of *implicitly defined by* on some definitions of that term.

<sup>6</sup>Assuming of course the proof system is sound with respect to the system of semantic constraints. In the case of semantic definability we do not need such a provision. Tarski cites a very near result, attributing it to Padoa: “In order [...] to show that a term ‘ $a$ ’ cannot be defined by means of the terms of a set  $B$  on the basis of a set  $X$  of sentences, it suffices to give two interpretations of all extra-logical constants which occur in the sentences of  $X$ , such that (1) in both interpretations all sentences of the set  $X$  are satisfied and (2) in both interpretations all the terms of the set  $B$  are given the same sense, but (3) the sense of the term ‘ $a$ ’ undergoes a change” ((Tarski, 1934), p. 300).

<sup>7</sup>Some comments are in place. 1. Compositionality is sometimes characterized as pertaining to *meanings*

*Remark.* Let  $L$  be a language that is compositional w.r.t. a set of semantic constraints  $\Delta$ . A term  $a$  in  $L$  is a logical term w.r.t.  $\Delta$  iff any phrase  $p$  consisting of the terms  $a_1, \dots, a_n$  and auxiliary symbols is determined by  $\{a_i : 1 \leq i \leq n, a_i \neq a\} \cup \{\mathcal{D}\}$ .

**Definition 12** (Extensional language) A language  $L$  is *extensional* (w.r.t.  $\Delta$ ) if every sentence is determined by the domain and the terms that constitute the sentence.<sup>8</sup>

Giving place to cases of less than full determination of the extension of terms, we may define notions of *dependency*, following which we define a *family* of terms.

**Definition 13** (Dependency) A set of phrases  $A$  *depends* on the set of phrases  $B$  (w.r.t.  $\Delta$ ) if there are  $\Delta$ -models  $M = \langle D, I \rangle$  and  $M' = \langle D, I' \rangle$  sharing the same domain  $D$  such that for any  $\Delta$ -model  $M^* = \langle D, I^* \rangle$ , if  $I^*(b) = I(b)$  for all  $b \in B$ , then  $I^*(a) \neq I'(a)$  for some  $a \in A$  (that is, fixing the phrases in  $B$  in a certain way excludes some interpretation for the phrases in  $A$  that can otherwise be realized).

A set of phrases  $A$  is *independent* of the set of terms  $B$  if it does not depend on it.

We say that a phrase  $a$  depends (is independent of) the set of phrases  $B$  if  $\{a\}$  depends on (is independent of)  $B$ .<sup>9</sup>

---

of terms. Here we have an extensional setting, where the only semantic value a term is given is its extension in a model (i.e. the value of  $I$ ). 2. In common notions of compositionality, structure plays a role as well, so that the value of a phrase is determined by its structure as well as its constituents. The framework we have here is quite simplified in that the structure of a phrase consists only of the order of its constituents and the position of auxiliary symbols such as parentheses. Note that the structure of a phrase does not vary across different models. Thus, in the present defined used of *determinateness* structure need not be mentioned.

<sup>8</sup>In the current extensional setting, extensionality is the particular case of compositionality concerning sentences. See previous footnote.

<sup>9</sup>Here we diverge from Tarski's notion of *independence*, which is weaker—for a term  $a$  to be independent of the terms in  $B$ , he requires only that the term  $a$  will not be definable by the terms in  $B$  (*ibid*). Tarski's definition of independence has the strange consequence that seemingly dependent terms (in the line of the examples given in this chapter) are ruled as independent if one does not completely determine the other. This consequence is avoided by my definition of independence.

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

In simpler words, the interpretations of dependent terms somehow restrict one another, while in the case of independent terms, all combinations of possible interpretations are realized in some model. The notion of “all possible combinations” can be represented using cartesian products. See below an alternative definition for dependency along those line.

So, for instance, by the constraint  $I(\textit{bachelor}) \subseteq I(\textit{unmarried})$ , *bachelor* depends on  $\{\textit{unmarried}\}$ : let  $I(\textit{unmarried}) = \{\textit{John}, \textit{Mary}\}$ ,  $I'(\textit{bachelor}) = \{\textit{John}, \textit{Jim}\}$ , so for any  $I^*$  such that  $I^*(\textit{unmarried}) = I(\textit{unmarried})$ ,  $I^*(\textit{bachelor}) \neq I'(\textit{bachelor})$ .

Note that if  $a$  is not a logical term and is determined by the terms in  $B$ , then it depends on the terms in  $B$ . This is proved below (Proposition 3).

Note also that dependency is a similarity relation, that is, it is reflexive and symmetric.

To end this subsection, we provide alternative definitions for *dependency* and *determinateness*, using new notation defined below. These definitions connect these concepts in a more perspicuous manner to the set-theoretic concepts of *cartesian product* (for dependency) and *function* (for determinateness).

Let

$$Ext_{\Delta}^n(D, a_1, \dots, a_n) := \{\langle I(a_1), \dots, I(a_n) \rangle : \langle D, I \rangle \text{ is a } \Delta\text{-model}\}$$

So, for instance,

$$Ext_{\Delta}^1(D, a) = \{I(a) : \langle D, I \rangle \text{ is a } \Delta\text{-model}\}$$

,

and

$$Ext_{\Delta}^2(D, a, b) := \{\langle I(a), I(b) \rangle : \langle D, I \rangle \text{ is a } \Delta\text{-model}\}$$

We omit the superscript  $n$ , since the arity of  $Ext$  is always determined by context.

Obviously,

$$Ext_{\Delta}(D, a, b) \subset Ext_{\Delta}(D, a) \times Ext_{\Delta}(D, b)$$

We can define *dependency* of  $a$  on  $b$  as:

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

There is a domain  $D$  such that

$$Ext_{\Delta}(D, a, b) \subsetneq Ext_{\Delta}(D, a) \times Ext_{\Delta}(D, b)$$

We can define the *determination* of  $b$  by  $a$  as:<sup>10</sup>

$$\bigcup_D Ext_{\Delta}(D, a, b) \text{ is a function.}$$

The reader can verify that these alternative definitions are equivalent to the original ones given for *determinateness* and *dependency* with respect to single terms.

For illustration, assume as before that  $\Delta$  includes the semantic constraint

$$I(allRed) \cap I(allGreen) = \emptyset.$$

According to Definition 13, *allRed* depends on *allGreen*. Indeed, take two  $\Delta$ -models:  $M = \langle D, I \rangle$  and  $M' = \langle D, I' \rangle$  sharing the same domain  $D$  such that  $I(allRed) = I'(allGreen) = C$ , where  $C$  is some non-empty subset of  $D$ . It is easy to see that there are such  $\Delta$ -models. Now, no  $\Delta$ -model is such that it interprets *allRed* as does  $M$  and *allGreen* as does  $M'$ , otherwise the constraint just formulated would be violated. This substantiates the dependence of *allRed* on *allGreen*. Note, also, using the notation just introduced, that

$$\langle C, C \rangle \in Ext_{\Delta}(D, allRed) \times Ext_{\Delta}(D, allGreen)$$

but

$$\langle C, C \rangle \notin Ext_{\Delta}(D, a, b).$$

---

<sup>10</sup>We mean *function* in a general sense, not restricted to sets (a collection of ordered pairs such that for each argument there is just one value), since the union on all domains is strictly speaking a proper class.

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

Similarly, we can extend the notation to deal with sets of terms.

For sets of terms  $A_1, \dots, A_n$ , let

$$Ext_{\Delta}^n(D, A_1, \dots, A_n) := \{ \{ \langle a, I(a) \rangle : a \in A_1 \}, \dots, \{ \langle a, I(a) \rangle : a \in A_n \} : \langle D, I \rangle \text{ is a } \Delta\text{-model} \}$$

Again, we omit the superscript, so that for a set of terms  $A$ :

$$Ext_{\Delta}(D, A) = \{ \{ \langle a, I(a) \rangle : a \in A \} : \langle D, I \rangle \text{ is a } \Delta\text{-model} \}$$

and for sets of terms  $A$  and  $B$ :

$$Ext_{\Delta}(D, A, B) = \{ \{ \langle a, I(a) \rangle : a \in A \}, \{ \langle b, I(b) \rangle : b \in B \} : \langle D, I \rangle \text{ is a } \Delta\text{-model} \}$$

Note that  $Ext_{\Delta}(D, a_1, \dots, a_n)$  and  $Ext_{\Delta}(D, A_1, \dots, A_n)$  are never empty when  $A_1, \dots, A_n$  are non-empty.

We note, similarly to before, that for any sets of terms  $A$  and  $B$ :

$$Ext_{\Delta}(D, A, B) \subset Ext_{\Delta}(D, A) \times Ext_{\Delta}(D, B).$$

We can define the *dependency* of  $A$  on  $B$ :

There is a domain  $D$  such that

$$Ext_{\Delta}(D, A, B) \subsetneq Ext_{\Delta}(D, A) \times Ext_{\Delta}(D, B)$$

We can define the *determination* of  $a$  by  $B$  as:

$$\bigcup_D \text{Ext}_\Delta(D, B, \{a\}) \text{ is a function.}$$

Again, the reader can verify that these alternative definitions are equivalent to the original ones given for *determinateness* and *dependency*.

Now we can easily see that:

**Proposition 3** *For every term  $a$  and set of terms  $B$ :*

1. *If  $a$  is a logical term, then  $a$  is independent of  $B$ .*
2. *If  $a$  is determined by  $B$ , and  $a$  is not a logical term, then  $a$  depends on  $B$ .*

*Proof.* 1. First we show that if  $a$  is a logical term, then it is independent of  $B$ . Assume that  $a$  is a logical term. Then for each domain  $D$ ,  $|\text{Ext}(D, a)| = 1$ . Now assume towards contradiction that there is some domain  $D_0$  such that

$$\text{Ext}_\Delta(D_0, B, \{a\}) \subsetneq \text{Ext}_\Delta(D_0, B) \times \text{Ext}_\Delta(D_0, \{a\}).$$

So there are  $I_1$  and  $I_2$  such that  $\langle D_0, I_1 \rangle$  and  $\langle D_0, I_2 \rangle$  are both  $\Delta$ -models such that

$$\langle \{ \langle b, I_2(b) \rangle : b \in B \}, \{ I_1(a) \} \rangle \notin \text{Ext}_\Delta(D_0, B, \{a\}).$$

Since  $|\text{Ext}(D_0, a)| = 1$ ,  $I_1(a) = I_2(a)$ . But then

$$\langle \{ \langle b, I_2(b) \rangle : b \in B \}, \{ I_1(a) \} \rangle = \langle \{ \langle b, I_2(b) \rangle : b \in B \}, \{ I_2(a) \} \rangle \in \text{Ext}_\Delta(D_0, B, \{a\})$$

and we reach contradiction.

2. Now assume that  $a$  is determined by  $B$  and that  $a$  is not a logical term. Then there is a domain  $D_0$  and there are  $I_1$  and  $I_2$  such that  $\langle D_0, I_1 \rangle$  and  $\langle D_0, I_2 \rangle$  are  $\Delta$ -models,

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

and such that  $I_1(a) \neq I_2(a)$ . By determination,  $\bigcup_D \text{Ext}_\Delta(D, B, \{a\})$  is a function.

So since

$$\langle \{ \langle b, I_1(b) \rangle \}, \{ I_1(a) \} \rangle \in \text{Ext}_\Delta(D_0, B, \{a\})$$

then

$$\langle \{ \langle b, I_1(b) \rangle \}, \{ I_1(a) \} \rangle \notin \text{Ext}_\Delta(D_0, B, \{a\}).$$

But

$$\langle \{ I_1(b) : b \in B \}, \{ I_2(a) \} \rangle \in \text{Ext}_\Delta(D, B) \times \text{Ext}_\Delta(D, \{a\})$$

so

$$\text{Ext}_\Delta(D_0, B, \{a\}) \subsetneq \text{Ext}_\Delta(D_0, B) \times \text{Ext}_\Delta(D_0, \{a\}).$$

and so  $a$  depends on  $B$ .

□

*Remark:* We have as a somewhat unnatural result that any logical term is independent of  $\{\mathcal{D}\}$ , although it is determined by it.

### 7.3.4 Categories and schemas

When taking all such constraints as those above to be determining the form of a sentence, it seems we might defy a basic intuition, whereby the form of a sentence can be represented by a *schema*. The problem, in essence, is the problem of substitution raised in Chapter 6.4. As was presented before, the problem was that the constraint  $I(\text{allRed}) \cap I(\text{allGreen}) = \emptyset$ , together with the standard constraints of first order logic, makes the sentence  $\varphi : \neg \exists x (\text{allRed}(x) \wedge \text{allGreen}(x))$  logically true, while its substitution instance  $\psi : \neg \exists x (\text{Giant}(x) \wedge \text{allGreen}(x))$  is not logically true. Thus, it would seem that  $\varphi$ 's form cannot be represented by a schema. This is given the basic assumptions that for any sentence  $\varphi$ , a substitution instance of  $\varphi$  is an instance of  $\varphi$ 's schema, and that all instances of a schema of a logically true sentence are logically true.



## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

The solution for the seeming failures of substitution and schematization is to diverge from the standard notions of substitution class and schematic letters. The idea is that terms are not divided into substitution classes according to grammatical categories, but according to their “behavior” in the system. Similarly, schematic letters will not be assigned to terms just by grammatical category.

Let us begin with an illustration of the proposed solution for schematization. In the case where we have a standard language augmented with semantic constraints for color terms, we may adjust the syntax and have a special set of schematic predicate symbols (metalinguistic variables) for color terms. So a sentence such as “There are no green giants” will be schematized as:  $\neg\exists x(\mathcal{G}x \wedge Gx)$  where upper-case letters in calligraphy range over color terms, and the rest of the symbols are as usual. So while the last schema is invalid in our system, the following one is valid:  $\neg\exists x(\mathcal{G}x \wedge \mathcal{R}x)$ .<sup>11</sup> I show below that in any system of semantic constraints it is possible in principle to represent all forms of sentences by schemas. To this end, I use the definitions that follow.

Let  $\pi$  be a permutation on a set  $A$  of terms of  $L$ .  $\pi$  can be extended to all terms of  $L$ , so that it is constant on all terms not in  $A$  (for any term of  $L$   $a$  such that  $a \notin A$ ,  $\pi(a) = a$ ).  $\pi$  can also be extended to the set of phrases of  $L$  in a natural way.  $\pi$  can be further extended to apply to models: for  $M = \langle D, I \rangle$ ,  $\pi(M) = \langle D, I^* \rangle$  where for each phrase  $s$ ,  $I^*(s) = I(\pi(s))$ . Let us now define the notion of a *category* of terms, which will eventually give us the proper substitution classes. This definition relies on the notion of *interchangeability*, defined first.

**Definition 14** (Interchangeability) Two terms  $a$  and  $b$  are *interchangeable* (w.r.t.  $\Delta$ ) if for any permutation  $\pi$  on  $\{a, b\}$  and  $\Delta$ -model  $M$ ,  $\pi(M)$  is a  $\Delta$ -model.

---

<sup>11</sup> We should add that we need to give up the convention that the same term may be substituted for two different schematic letters (given they are of the right category). Otherwise the sentence above is comes out invalid when *allRed* is substituted for both schematic letters. This unconventional idea of schematization and identity is discussed in (Smiley, 1982; Wehmeier, 2004, 2012).

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

Interchangeability of terms with respect to a set of semantic constraints is an equivalence relation. So for any set  $\Delta$  of semantic constraints for a language  $L$  there is a partition of the terms of  $L$  into maximal interchangeability classes. Those will be the  $\Delta$ -categories:

**Definition 15** (Category) A set of terms  $A$  is a *category* (w.r.t.  $\Delta$ ) if every two terms in  $A$  are interchangeable, and no term  $a \in A$  is interchangeable with a term  $b \notin A$ .<sup>12</sup>

An example of a category is that of color terms in the set of constraints presented above. Note that the categories here are “semantic” in their nature, and do not in general coincide with what are usually thought of as syntactic or grammatical categories. In a system with constraints for color terms as the one above, “predicate” is not a category by our definition, since not all predicates behave alike in the system (e.g. *big* and *allRed* behave differently).

On the basis of the above definitions, we can now define a *schema*. The basic idea is that each category will have its own pool of schematic letters. Each schema has as instances sentences that are received by substituting a term from the right category for each schematic letter. To return to the example with which we started, assume we have a system of constraints where color terms form a category, and all other predicates form a category. The sentence  $\neg\exists x(Green(x) \wedge Giant(x))$  is then an instance of the schema  $\neg\exists x(\mathcal{G}x \wedge Gx)$  but not of the schema  $\neg\exists x(\mathcal{G}x \wedge \mathcal{R}x)$ , where  $\mathcal{G}$  and  $\mathcal{R}$  are schematic letters for color terms and  $G$  is a schematic term for the category of all other predicates.<sup>13</sup>

In order to give a more rigorous definition of *schema* in our system, we first define a *schematic phrase* and an *instance*. For the sake of these definitions, we make use of an

---

<sup>12</sup>The concept of *semantical category*, to which my concept of *category* is akin, is usually assumed and not defined. Tarski gives a strictly weaker definition of *semantical category* than the one I provide above, though the two definitions coincide in the case of nonlogical terms of standard languages (Tarski, 1933, pp. 215-217).

<sup>13</sup>Terms such as  $\neg$ ,  $\exists$  and  $\wedge$  can be used as schematic letters themselves where (as is in the standard setting) each of their categories is a singleton.

infinite sequence pairwise disjoint infinite sets of schematic letters, signified by  $\{S_1, S_2, \dots\}$ .

**Definition 16** (Instance) Let  $\Delta$  be a set of constraints, and  $T = \{T_1, T_2, \dots\}$  be a partition of the terms of  $L$  into categories for  $\Delta$  (which can be viewed as *substitution classes* for the logic). Let  $p = a_1 \dots a_n$  be a string of terms in  $L$  and auxiliary symbols (e.g. parentheses) and let  $\Phi = A_1 \dots A_k$  be a string of schematic letters and auxiliary symbols.  $p$  is an *instance* of  $\Phi$  (w.r.t.  $\langle L, \Delta, T \rangle$ ) if the following two conditions hold:

1.  $n = k$ .
2. For all  $i, j \in \{1, \dots, n\}$ :
  - i)  $A_i \in S_k \Leftrightarrow a_i \in T_k$ .
  - ii)  $A_i = A_j \Leftrightarrow a_i = a_j$ .
  - iii) If  $a_i$  is an auxiliary symbol, then  $A_i = a_i$ .

**Definition 17** (Schema) Let  $\Delta$  be a set of constraints, and  $T = \{T_1, T_2, \dots\}$  be a partition of the terms of  $L$  into categories for  $\Delta$ . Let  $\Phi = A_1 \dots A_k$  be a string of schematic letters and auxiliary symbols.  $\Phi$  is a *schema* (w.r.t.  $\langle L, \Delta, T \rangle$ ) if there is a sentence  $\varphi$  in  $L$  such that  $\varphi$  is an instance of  $\Phi$ .

**Definition 18** (Validity of schemas) A schema is *valid* if all its instances are valid. A schema is *invalid* if all its instances are not valid.

The validity of sentences is defined as usual: a sentence is *valid* if true in all models.

Validity is always relative to a language  $L$  and a set of constraints  $\Delta$  and a partition of the terms of  $L$  into  $\Delta$ -categories, which are left unmentioned if allowed by context.

**Proposition 4** *Every schema is either valid or invalid.*<sup>14</sup>

---

<sup>14</sup>Similarly, for any schema, either all its instances are *invalid* (i.e. logical falsehoods) or none are. The proof is similar to the one that follows in the main text. We can thus divide all schemas to those for which all instances are valid (i.e. logical truths), those for which all instances are invalid (i.e. logical falsehoods)

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

*Proof.* Let  $\Phi$  be a schema in  $\langle L, \Delta, T \rangle$ . Let  $\varphi$  and  $\psi$  be two instances. Let  $\varphi = a_1 \dots a_n$ ,  $\psi = b_1 \dots b_n$  (they are of the same length as  $\Phi$  so their lengths are equal).

First assume that  $\varphi$  and  $\psi$  have no terms in common. Assume toward contradiction that  $\varphi$  is valid and  $\psi$  isn't. Then there is a  $\Delta$ -model  $M = \langle D, I \rangle$  such that  $M \not\models \psi$ . We know that for all  $i$ , if  $a_i$  and  $b_i$  are terms (as opposed to auxiliary symbols), then they are of the same category, so they are interchangeable in  $\Delta$ .

Let  $\pi$  be a permutation over the terms in  $L$  such that for all  $i \in \{1, \dots, n\}$ ,  $\pi(a_i) = b_i$  and  $\pi(b_i) = a_i$  (where those are terms), and for any other term  $a$ ,  $\pi(a) = a$ .  $\pi$  can be extended to all symbols in  $\varphi$  and  $\psi$ , so that it will be constant on auxiliary symbols, and we have  $\pi(a_i) = b_i$  and  $\pi(b_i) = a_i$  for all  $i \in \{1, \dots, n\}$ . Since (when restricted to terms)  $\pi$  is the result of a finite composition of permutations (that are “switchings”), each restricted to a category,  $\pi(M)$  is a  $\Delta$ -model. However,  $\pi(M) \not\models \varphi$ : note that  $\pi(\varphi) = \pi(a_1) \dots \pi(a_n) = b_1 \dots b_n = \psi$  so  $\pi(M) \models \varphi$  iff  $M \models \psi$ . But  $M \not\models \psi$ , so  $\pi(M) \not\models \varphi$ , and from this follows a contradiction to our assumption that  $\varphi$  is valid in  $\Delta$ .

Assume now that  $\varphi$  and  $\psi$  have some terms in common: for some  $i, j \in \{1, \dots, n\}$ ,  $a_i = b_j$ . Assume that  $\varphi$  is valid. For each  $i \in \{1, \dots, n\}$  such that  $b_i$  is a term, add to  $L$  the new term  $b'_i$  such that  $b'_i = b'_j$  iff  $b_i = b_j$ . For every phrase  $p$  in  $L$  where  $b_i$  occurs for some  $i$ , add the phrase  $p'$  which is received from  $p$  by replacing all  $b_i$ s with corresponding  $b'_i$ s. Add to  $\Delta$  the constraint  $I(p') = I(p)$  for every phrase  $p$  in  $L$  and every phrase  $p'$  obtained by replacing some  $b_i$ s with corresponding  $b'_i$ s. Call the new language  $L'$  and the new set of constraints  $\Delta'$ . Note that the extension made is conservative in the sense that every sentence  $\chi$  with terms only from  $L$  is valid in  $\langle L', \Delta' \rangle$  iff it is valid in  $\langle L, \Delta \rangle$ . Obviously,

---

and those for which all instances are neither valid nor invalid. The basic idea of the proof is that for every two instances of a schema in  $\langle L, \Delta \rangle$  and a  $\Delta$ -model for one of the instances there is a  $\Delta$ -model for the other. Note that the notion of *schema* defined here diverges from the standard notion, by which invalid schemas may have a valid instance. On this account, I take it that my notion tracks better the notion of *logical form*. See also f.n. 11.

each  $b'_i$  is in the same  $\Delta'$ -category as  $b_i$ , and other than the addition of the  $b_i$ s the categories in  $\Delta'$  are the same as in  $\Delta$ .<sup>15</sup>

For each  $i$  such that  $b_i$  is an auxiliary symbol, let  $b'_i = b_i$ . Now let  $\psi' = b'_1 \dots b'_n$ . The process of schematization can be done with respect to  $L'$  and  $\Delta'$ , and it can be easily verified that there is a schema of which  $\psi'$ ,  $\psi$  and  $\varphi$  are instances. Since  $\varphi$  is valid in  $\langle L, \Delta \rangle$ , then  $\varphi$  is valid in  $\langle L', \Delta' \rangle$ .  $\psi'$  and  $\varphi$  have no terms in common. So, according to what was proved before,  $\psi'$  is valid in  $\langle L', \Delta' \rangle$ . Also, clearly, for any  $\Delta'$ -model  $M$  for  $L'$ ,  $M \models \psi'$  iff  $M \models \psi$ . So  $\psi$  is valid in  $\langle L', \Delta' \rangle$ . Because the extension of the language was conservative as explained above, it follows that  $\psi$  is valid in  $\langle L, \Delta \rangle$ .  $\square$

Proposition 4 shows that schemas can serve as or represent forms of sentences: a sentence will be valid iff all instances of its schema are valid. A sentence will be nonvalid iff all instances of its schema are nonvalid. In a similar manner, we can schematize *arguments*, and define validity for argument-schemas. It will then follow that an argument is valid iff its argument-schema is valid.

However, in many cases the use of schemas may be impractical. We may have a very large number of categories, and in cases in which most categories are singletons, most schemas will have only one instance and there will be no benefit from using them. Extending the notions of category and schema to deal with *families* of terms may be more useful. I define the generalized notions of a category and of a schema in what follows.

---

<sup>15</sup>We need to add here the assumption that semantic constraints are enough well-behaved in the sense that terms can be added to the language without canceling any models. That is, if  $M$  is a  $\Delta$ -model for  $L$ ,  $b'$  a term in  $L'$  but not in  $L$ , and  $M'$  a model for  $L'$  that agrees with  $M$  on all phrases in  $L$ , then  $M'$  is a  $\Delta$ -model for  $L'$ . This assumption holds for all “reasonable” sets of semantic constraints, including all examples given in this chapter.

### 7.3.5 Families of terms

Let  $Terms$  be the set of terms in  $L$ . Assume  $T$  is a partition of  $Terms$ , that is:  $T$  is a set of pairwise disjoint subsets of  $Terms$ , and  $\cup T = Terms$ . Assume also that each set in  $T$  is ordered. Call such a sequence of terms a *family*. Families can be either finite or infinite.

**Definition 19** (Maximal family) A family  $(a_1, \dots)$  (either finite or infinite) is a *maximal family* (*maxfam*) if the set of its members  $\{a_i\}_i$  is independent of  $Terms \setminus \{a_i\}_i$ .

**Definition 20** (Isomorphism of families) Let  $\mathcal{F}_1$  and  $\mathcal{F}_2$  be two equinumerous families, that is:  $\mathcal{F}_1 = (a_1, \dots, a_n)$  and  $\mathcal{F}_2 = (b_1, \dots, b_n)$  for some natural  $n$ , or both  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are infinite.  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are *isomorphic* (w.r.t.  $\Delta$ ) ( $\mathcal{F}_1 \cong^\Delta \mathcal{F}_2$ ) if they are interchangeable in  $\Delta$ . Namely, let  $\pi$  be the permutation on the terms of  $L$  such that for all  $i$ ,  $\pi(a_i) = b_i$  and  $\pi(b_i) = a_i$ , and for all  $a \notin \mathcal{F}_1 \cup \mathcal{F}_2$ ,  $\pi(a) = a$ , then for every  $\Delta$ -model  $M$ ,  $\pi(M)$  is a  $\Delta$ -model.

**Definition 21** (Tight (loose) family) Let  $\mathcal{F}_1$  and  $\mathcal{F}_2$  be two equinumerous families, that is:  $\mathcal{F}_1 = (a_1, \dots, a_n)$  and  $\mathcal{F}_2 = (b_1, \dots, b_n)$  for some natural  $n$ , or both  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are infinite.  $\mathcal{F}_1$  is *tighter* than  $\mathcal{F}_2$  ( $\mathcal{F}_2$  *looser* than  $\mathcal{F}_1$ ) (w.r.t.  $\Delta$ ) if for every  $\Delta$ -model  $M = \langle D, I \rangle$  there is a  $\Delta$ -model  $M' = \langle D, I' \rangle$  such that for all  $i$ ,  $I(a_i) = I'(b_i)$ .

**Proposition 5** Let  $\mathcal{F}_1$  and  $\mathcal{F}_2$  be two equinumerous families.

1. If  $\mathcal{F}_1 \cong \mathcal{F}_2$  then  $\mathcal{F}_1$  is both tighter and looser than  $\mathcal{F}_2$ .
2. If  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are maxfams, their underlying sets of terms are disjoint, and  $\mathcal{F}_1$  is both tighter and looser than  $\mathcal{F}_2$ , then  $\mathcal{F}_1 \cong \mathcal{F}_2$ .

*Proof.* 1. Assume  $\mathcal{F}_1 \cong \mathcal{F}_2$ . Let  $M = \langle D, I \rangle$  be a  $\Delta$ -model. Let  $\pi$  be as in Definition 20. Then  $\pi(M)$  is a  $\Delta$ -model.  $\pi(M) = \langle D, I^\pi \rangle$  is such that for all  $i$ ,  $I^\pi(b_i) = I(\pi(b_i)) = I(a_i)$  (since  $\pi$  switches  $a_i$  and  $b_i$ ), which is what we need. So  $\mathcal{F}_1$  is tighter than  $\mathcal{F}_2$ . Analogously,

$\mathcal{F}_1$  is looser than  $\mathcal{F}_2$ .

2. Assume  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are maxfams, and that  $\mathcal{F}_1$  is both tighter and looser than  $\mathcal{F}_2$ . Let  $\pi$  be as before. Let  $M = \langle D, I \rangle$  be a  $\Delta$ -model. Then there is a  $\Delta$ -model  $M' = \langle D, I' \rangle$  such that for all  $i$ ,  $I(a_i) = I'(b_i)$ , and there is a  $\Delta$ -model  $M'' = \langle D, I'' \rangle$  such that for all  $i$ ,  $I(b_i) = I''(a_i)$ . Note that  $\pi(M)$  agrees with  $M'$  on  $\mathcal{F}_2$  and with  $M''$  on  $\mathcal{F}_1$ .

Since  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are independent (since they are disjoint maxfams), there is a  $\Delta$ -model  $M'''$  that agrees with  $M'$  on  $\mathcal{F}_2$  and with  $M''$  on  $\mathcal{F}_1$ .

*Claim:*  $\mathcal{F}_1 \cup \mathcal{F}_2$  is independent of  $(\mathcal{F}_1 \cup \mathcal{F}_2)^C$ .

*Proof of the claim:* Let  $M_0 = \langle D_0, I_0 \rangle$  and  $M_1 = \langle D_1, I_1 \rangle$  be  $\Delta$ -models such that  $D_0 = D_1$ . We show that there is a  $\Delta$ -model  $M_* = \langle D_*, I_* \rangle$  such that  $D_* = D_0$  that agrees with  $M_0$  on  $\mathcal{F}_1 \cup \mathcal{F}_2$  and with  $M_1$  on  $(\mathcal{F}_1 \cup \mathcal{F}_2)^C$ .

Now there is a  $\Delta$ -model  $M_2$  that agrees with  $M_0$  on  $\mathcal{F}_1$  and with  $M_1$  on  $(\mathcal{F}_1 \cup \mathcal{F}_2)^C$  ( $\mathcal{F}_1$  is a maxfam). And there is a  $\Delta$ -model  $M_3$  that agrees with  $M_2$  on  $\mathcal{F}_1 \cup (\mathcal{F}_1 \cup \mathcal{F}_2)^C$  and with  $M_0$  on  $\mathcal{F}_2$  ( $\mathcal{F}_2$  is a maxfam), which is what we need:  $M_3$  agrees with  $M_0$  on  $\mathcal{F}_1$  and on  $\mathcal{F}_2$  and with  $M_1$  on  $(\mathcal{F}_1 \cup \mathcal{F}_2)^C$ , and this proves the claim.

Now there is a  $\Delta$ -model that agrees with  $M'''$  on  $\mathcal{F}_1 \cup \mathcal{F}_2$  and with  $M$  on  $(\mathcal{F}_1 \cup \mathcal{F}_2)^C$  which is precisely  $\pi(M)$  – so  $\pi(M)$  is a  $\Delta$ -model.  $\square$

Each family has an equivalence class with respect to interchangeability (whose members are all the families to whom it is isomorphic, see Definition 20). Such an equivalence class will be called a *(generalized) category*. Due to isomorphism, the length all families in a given generalized category is the same. Assign to each generalized category a set of *complex schematic variables*, where each is a sequence of schematic letters  $(\mathcal{S}_1, \dots)$ , either finite or infinite: of the same length as that of the families in the generalized category.

On the basis of these assumptions, we can define a *generalized schema*:

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

**Definition 22** (Generalized schema) Let a *schema* in  $\langle L, \Delta \rangle$  be a string of schematic letters and auxiliary symbols that has an instance.  $\varphi$  is an *instance* of  $\Phi$  if the following holds:

1.  $\varphi$  is a sentence.
2.  $\Phi$  and  $\varphi$  are of the same length (including auxiliary symbols in both).
3. Let  $\Phi = A_1 \dots A_n$  and  $\varphi = a_1 \dots a_n$ . For all  $i \in \{1, \dots, n\}$ :
  - i) If  $a_i$  is a term and is in the  $k$ -th place in a family  $\mathcal{F}$ , then  $A_i$  is in the  $k$ -th place in some complex schematic variable assigned to the category of  $\mathcal{F}$ .
  - ii) If  $a_i$  and  $a_j$  are terms, then they belong to the same family iff  $A_i$  and  $A_j$  belong to the same complex schematic variable.
  - iii) If  $a_i$  is an auxiliary symbol, then  $A_i = a_i$ .

As before, a generalized schema is  $\Delta$ -*valid* if all its instances are  $\Delta$ -valid (true in all  $\Delta$ -models). A schema is  $\Delta$ -*invalid* if none of its instances are  $\Delta$ -valid.

**Proposition 6** (*Generalized schematization*) *Each schema is either valid or invalid.*

*Proof.* Let  $\varphi$  and  $\psi$  be instances of  $\Phi$ . Assume toward contradiction that  $\varphi$  is valid and  $\psi$  is not valid. Let  $\varphi = a_1 \dots a_n$  and  $\psi = b_1 \dots b_n$  (they are of the same length because they are both the length of  $\Phi$ ). So there is a  $\Delta$ -model  $M$  such that  $M \not\models \psi$ .

First assume that  $\varphi$  and  $\psi$  have no terms in common, and moreover, that no term in  $\varphi$  belongs to the same family as a term in  $\psi$ . Let  $\pi$  be a permutation on *Terms* such that for all  $i \in \{1, \dots, n\}$ ,  $\pi(a_i) = b_i$  and  $\pi(b_i) = a_i$  where those are terms (as opposed to auxiliary symbols), and  $\pi$  is the identity on all other terms in  $L$ .  $\pi$  can be extended to all symbols in  $\varphi$  and  $\psi$ , so that it will be constant on auxiliary symbols, and we have  $\pi(a_i) = b_i$  and  $\pi(b_i) = a_i$  for all  $i \in \{1, \dots, n\}$ .



*Claim:* there is a series of permutations  $(\pi_1, \dots, \pi_k)$  such that  $\pi = \pi_k \cdots \pi_1$  and each of the permutation in the series is a permutation that interchanges two families.

*Proof of the claim:* For each  $i \in \{1, \dots, n\}$  let  $\mathcal{F}_i^\phi$  be the family to which  $a_i$  belongs and  $\mathcal{F}_i^\psi$  the family to which  $b_i$  belongs.  $\mathcal{F}_i^\phi$  and  $\mathcal{F}_i^\psi$  are of the same category (by the definition of a schema), and so are isomorphic. Define  $\pi'_i$  to be the permutation that interchanges the two families. For all  $i, j \in \{1, \dots, n\}$ ,  $a_i$  is in the same family of  $a_j$  iff  $b_i$  is in the same family of  $b_j$ , so  $\pi'_i$  and  $\pi'_j$  either interchange the same families or the two pairs of families they interchange do not coincide. The permutations that we finally take, the  $\pi_i$ 's, will be the  $\pi'_i$ 's without repetitions. From the construction of the  $\pi_i$ 's it follows that  $\pi = \pi_n \cdots \pi_1$ .

$\pi_1(M)$  is a  $\Delta$ -model, and by direct induction  $\pi(M)$  is a  $\Delta$ -model. However,  $\pi(M) \not\models \varphi$ : note that  $\varphi = a_1 \dots a_n = \pi(b_1) \dots \pi(b_n) = \pi(\psi)$  so  $\pi(M) \models \varphi$  iff  $M \models \psi$ . But  $M \not\models \psi$ , so  $\pi(M) \not\models \varphi$ , and from this follows a contradiction to our assumption that  $\varphi$  is valid in  $\Delta$ .

Now assume that  $\varphi$  and  $\psi$  have terms that belong to the same family. The proposition then follows by an analogous argument to the one in the proof of Proposition 4, through the extension of the language and set of constraints.

□

### 7.3.6 Families: Examples

Let  $L$  be a language. We assume that  $L$  contains the following unary predicates: *red*, *green*, *blue*, *small*, *medium*, *big* and *extended*, and the binary predicate *bigger*. Other than that,  $L$  includes the usual first order vocabulary. We formulate the following constraints for  $L$ :

1.  $I(\text{red}) \cap I(\text{green}) = \emptyset$
2.  $I(\text{red}) \cap I(\text{blue}) = \emptyset$
3.  $I(\text{blue}) \cap I(\text{green}) = \emptyset$

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

4.  $I(\text{small}) \cap I(\text{medium}) = \emptyset$
5.  $I(\text{small}) \cap I(\text{big}) = \emptyset$
6.  $I(\text{big}) \cap I(\text{medium}) = \emptyset$
7.  $I(\text{red}) \cup I(\text{green}) \cup I(\text{blue}) \subseteq I(\text{extended})$
8.  $I(\text{small}) \cup I(\text{medium}) \cup I(\text{big}) \subseteq I(\text{extended})$
9.  $I(\text{bigger}) \subseteq I(\text{extended}) \times I(\text{extended})$
10.  $((I(\text{big}) \times I(\text{medium})) \cup (I(\text{big}) \times I(\text{small})) \cup (I(\text{medium}) \times I(\text{small}))) \subseteq I(\text{bigger})$

Let  $\Delta$  be the set of semantic constraints for  $L$  including the standard constraints for first order logic (see Section 7.2.2), with the addition of constraints 1-6. Let  $\Delta^+$  be  $\Delta$  with the addition of constraints 7-8, and let  $\Delta^*$  be  $\Delta^+$  with the addition of constraints 9-10.

We define the following families:

$$\mathcal{F}_1 = \langle \text{red}, \text{green}, \text{blue} \rangle$$

$$\mathcal{F}_2 = \langle \text{small}, \text{medium}, \text{big} \rangle$$

$$\mathcal{F}_1^+ = \langle \text{red}, \text{green}, \text{blue}, \text{extended} \rangle$$

$$\mathcal{F}_2^+ = \langle \text{small}, \text{medium}, \text{big}, \text{extended} \rangle$$

$$\mathcal{F}_1^* = \langle \text{red}, \text{green}, \text{blue}, \text{extended}, \text{bigger} \rangle$$

$$\mathcal{F}_2^* = \langle \text{small}, \text{medium}, \text{big}, \text{extended}, \text{bigger} \rangle$$

### Facts

- i.  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are maxfams in  $\Delta$  (see Definition 19, p. 145), but not in  $\Delta^+$  and  $\Delta^*$ .
- ii.  $\mathcal{F}_1 \cup \mathcal{F}_2$  is a maxfam in  $\Delta$ , but not in  $\Delta^+$  and  $\Delta^*$ .  $\mathcal{F}_1^+ \cup \mathcal{F}_2^+$  is a maxfam in  $\Delta$  and  $\Delta^+$ , but not in  $\Delta^*$ .  $\mathcal{F}_1^* \cup \mathcal{F}_2^*$  is a maxfam in  $\Delta$ ,  $\Delta^+$  and  $\Delta^*$ .
- iii.  $\mathcal{F}_1 \cong^\Delta \mathcal{F}_2$ ,  $\mathcal{F}_1 \cong^{\Delta^+} \mathcal{F}_2$ , but  $\mathcal{F}_1 \not\cong^{\Delta^*} \mathcal{F}_2$ . Similarly,  $\mathcal{F}_1^+ \cong^\Delta \mathcal{F}_2^+$ ,  $\mathcal{F}_1^+ \cong^{\Delta^+} \mathcal{F}_2^+$ , but  $\mathcal{F}_1^+ \not\cong^{\Delta^*} \mathcal{F}_2^+$ . And similarly,  $\mathcal{F}_1^* \cong^\Delta \mathcal{F}_2^*$ ,  $\mathcal{F}_1^* \cong^{\Delta^+} \mathcal{F}_2^*$ , but  $\mathcal{F}_1^* \not\cong^{\Delta^*} \mathcal{F}_2^*$ .
- iv.  $\mathcal{F}_1$  is both tighter and looser than  $\mathcal{F}_2$  in  $\Delta$  and  $\Delta^+$ , but not in  $\Delta^*$ , where  $\mathcal{F}_1$  is looser but not tighter than  $\mathcal{F}_2$ .

Finally, we show an example for a generalized schematization in  $\Delta^+$ . First, we need to list all the categories with respect to  $\Delta^+$ . Most of the categories consist of singleton families. We have one category which consists of two families, each a sequence of three members. Those are:

- Distinct categories for the standard quantifiers and connectives:  $\{< \neg >\}$ ,  $\{< \rightarrow >\}$ ,  $\{< \wedge >\}$ ,  $\{< \vee >\}$ ,  $\{< \forall >\}$  and  $\{< \exists >\}$ .
- The category of the family of color terms and the family of size terms:  $\{< red, green, blue >, < small, medium, big >\}$
- The singleton category  $\{< extended >\}$
- For each  $n$ , the category  $\{< P >: P \text{ is an } n\text{-ary predicate, } P \notin \{red, green, blue, small, medium, big, extended\}\}$
- The category of individual constants.<sup>16</sup>

Secondly, to construct schemas, we need a pool of schematic letters for each category.

For any singleton category, we can simply use the term  $\backslash s$  in that category to stand as

---

<sup>16</sup>As before, we leave out function symbols.

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

schematic letters themselves. We shall use a dot above the symbol to indicate that it is a schematic letter rather than a term in the language. Thus, the schematic letter for  $\vee$  will be  $\dot{\vee}$ .

We shall use the following schematic letters for the family of triples above:

$(A_1^1, A_1^2, A_1^3), (A_2^1, A_2^2, A_2^3), (A_3^1, A_3^2, A_3^3), (A_4^1, A_4^2, A_4^3) \dots$  (The way these schematic letters are used will be demonstrated below.)

Finally, we have schematic letters for the category of unconstrained predicates:  $\mathfrak{P}_1, \mathfrak{P}_2, \dots$  and individual constants:  $\mathfrak{a}_1, \mathfrak{a}_2, \dots$

Auxiliary symbols, such as comma and parentheses, are not altered in the process of schematization. We treat variables as punctuation marks.

We now list some schemas, their logical status in  $\Delta^+$ , and some of their instances.

- $\mathfrak{P}_1(\mathfrak{a}_1)$ , invalid,  
*naturalNumber(john), bachelor(a).*
- $A_1^1(\mathfrak{a}_1)$ , invalid,  
*red(a), small(b).*
- $A_1^2(\mathfrak{a}_1) \dot{\rightarrow} \text{extended}(\mathfrak{a}_1)$ , valid,  
*green(a)  $\rightarrow$  extended(a), medium(john)  $\rightarrow$  extended(john).*
- $\dot{\wedge}(A_1^1(\mathfrak{a}_1) \dot{\wedge} A_1^2(\mathfrak{a}_1))$ , valid,  
 *$\neg(\text{red}(a) \wedge \text{green}(a)), \neg(\text{small}(b) \wedge \text{medium}(b)).$*
- $A_1^1(\mathfrak{a}_1) \dot{\wedge} A_2^1(\mathfrak{a}_1)$ , invalid,  
 *$\text{red}(a) \wedge \text{small}(a), \text{small}(a) \wedge \text{red}(a).$*
- $\mathfrak{P}_1(\mathfrak{a}_1) \dot{\rightarrow} \dot{\wedge} A_1^3(\mathfrak{a}_1)$ , invalid,  
 *$\text{naturalNumber}(a) \rightarrow \neg \text{blue}(a), \text{bachelor}(\text{john}) \rightarrow \neg \text{big}(\text{john})$*

### 7.3.7 Equivalence

**Definition 23** (Equivalence) Two sets of semantic constraints  $\Gamma$  and  $\Delta$  to the same language  $L$  are *equivalent*, written  $\Gamma \equiv \Delta$ , if  $\Gamma$  and  $\Delta$  have the same range of admissible models (that is, every  $\Gamma$ -model is a  $\Delta$ -model and vice versa).

In Chapter 6 I have proposed that the collection of semantic constraints represents, in a very general sense, the form, or structure, of the language. If so, shall we say that equivalent sets of constraints give rise to, or represent, the same form? Equivalent sets of constraints to the same language give identical logical consequence relations. Moreover, in the definitions given in this chapter, no significance has been given to the differences between equivalent sets of constraints: we may have just as well concentrated on classes of models rather than with the sets of constraints defining them. Yet I would like to maintain the possibility of a more intensional outlook, where the formulation of the semantic constraints is significant. This might especially be appropriate if we conceive of semantic constraints as *explicit commitments* we undertake with respect to language, as suggested in Chapter 10.2. Thus, at this point, although the framework presented invites an extensional view by which equivalent sets of constraints give rise to the same form, I would like to leave this issue open.

## 7.4 A Remark on Proof Theory

The framework of semantic constraints, as I have presented it, is founded on model-theoretic semantics. It can be worthwhile to note that the basic ideas underlying the system can be spelled out in alternative ways. In particular, some of these ideas can be translated to a proof-theoretic setting. The upshot will be having rules that constrain the behavior of terms without necessarily “fixing” them completely. Instead of *semantic constraints*, we

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

can speak of *constraining rules*.<sup>17</sup> Here we shall consider an easy case of constraining rules. The point can be made using the following examples (for which we assume an FOL setting, that  $t$  is a singular term, and that  $\perp$  is a term in L):

$$R_1 \quad \frac{red(t) \quad green(t)}{\perp}$$

$$R_2 \quad \frac{bachelor(t)}{unmarried(t)}$$

$R_1$  and  $R_2$  can be incorporated into a natural deduction system. The rules  $R_1$  and  $R_2$  add some inferential information regarding *red*, *green*, *bachelor* and *unmarried*. This information does not yield a characterization of these terms through introduction and elimination rules, but constrains them in some manner. Through such rules, we re-articulate the idea that a logical system need not be limited to the strict categories of “logical” and “nonlogical” terms.

In this work, I shall not explore the avenue of proof-theoretic semantic constraints. But some obvious results can be mentioned. Assume L is a first order language in the standard sense. Assume a standard natural deduction system for L, augmented with  $R_1$  and  $R_2$ . On the model-theoretic side, assume a system of constraints for first order logic (as presented in 7.2.2) augmented with the constraints:

$$C_1 \quad I(red) \cap I(green) = \emptyset$$

$$C_2 \quad I(bachelor) \subseteq I(unmarried)$$

One can see that the proof-theoretic system matches the model-theoretic one. That is, it is sound and complete. Soundness can be proved similar to the standard way by induction

---

<sup>17</sup>This is to avoid the confusion that might arise from the use of *semantics* as pertaining to the model-theoretic tradition as opposed to the proof-theoretic one — although the notion of *semantics* is employed in some approaches to proof theory as well.

on proofs.<sup>18</sup>

We can prove soundness and completeness fairly easily using standard soundness and completeness results, as  $C_1$  and  $C_2$  can be expressed by sentences in  $L$ :

$$A_1 \quad \neg \exists x (red(x) \wedge green(x))$$

$$A_2 \quad \forall x (bachelor(x) \rightarrow unmarried(x))$$

Let  $FOL$  be the standard set of semantic constraints for first order logic. Let  $\Delta$  be the set of semantic constraints including  $FOL$  and  $C_1$  and  $C_2$ . It is easy to see that  $M$  is a  $\Delta$ -model if and only if  $M$  is an  $FOL$ -model such that  $M \models A_1$  and  $M \models A_2$ . We use ‘ $\vdash_{FOL}$ ’ as the provability sign for the standard proof-system for first order logic.  $R_1$ ,  $R_2$  and the standard rules for first order logic are part of the proof-system for  $\Delta$ , and ‘ $\vdash_\Delta$ ’ is used as the corresponding provability sign.

**Proposition 7** (Soundness and completeness for  $\Delta$ ) *Let  $\Gamma$  be a set of sentences in  $L$ , and  $\varphi$  a sentence in  $L$ . Then  $\Gamma \models_\Delta \varphi$  iff  $\Gamma \vdash_\Delta \varphi$ .*

*Proof.* The remarks above tell us that  $\Gamma \models_\Delta \varphi$  iff  $\Gamma \cup \{A_1, A_2\} \models_{FOL} \varphi$ . Now by soundness and completeness for  $FOL$ ,  $\Gamma \cup \{A_1, A_2\} \models_{FOL} \varphi$  iff  $\Gamma \cup \{A_1, A_2\} \vdash_{FOL} \varphi$ . We’re done once we show that  $\Gamma \cup \{A_1, A_2\} \vdash_{FOL} \varphi$  iff  $\Gamma \vdash_\Delta \varphi$ . Left to right: it suffices to show that  $\vdash_\Delta A_1$  and  $\vdash_\Delta A_2$ . The derivations are straightforward. Right to left: it suffices to show that  $R_1$  and  $R_2$  are derived rules in  $FOL$  given  $A_1$  and  $A_2$  as assumptions. Here, too, the derivations are straightforward.  $\square$

Notably, the possibility of having a complete proof system for a set of constraints depends on the expressive power of the object language with respect to those constraints.

---

<sup>18</sup>Remember that I am assuming here a “normal” syntax. The only change in the regular manner of proof is that there will be additional cases for the special predicates we singled out. A diversion from standard syntax (for instance, with different or less strict grammatical categories, see Section 7.2.3 for examples) will not exclude proof by induction on formulas, but just change its nature and its power. I will not get into this issue here for reasons of focus and scope.

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

A constraint that fixes the extension of the term *naturalNumber* will not have a rule correlate, nor a corresponding (effective) system of rules, on pain of incompleteness. Of course, this fact can be used not just to show the limitations of proof theory, but also in an argument against having such a constraint. Having constraints that are expressible by rules or axioms in the object language has the obvious benefit of manageability in a proof system.

In any case, regardless of completeness issues, if we endorse a proof-theoretic perspective, taking rules as primary, we are limited to the expressive power of the object language. This is still consistent with the ideology underlying the framework of semantic constraints. As mentioned above, the rules version as the model-theoretic version dismisses the centrality of logical terms in formality and logic. As will be elaborated in Chapter 10.2, semantic constraints can be viewed as commitments made in a context where reasoning is the main focus. These commitments can also take the form of explicitly set rules. However, the orientation of this work is model-theoretic, and with these remarks I conclude the detour into proof theory.

### 7.5 Appendix: p-categories

In Section 7.3.4 categories were defined through the notion of interchangeability: a category is a maximal set of interchangeable terms. Here I would like to explore another, more general notion of category, where any subset of terms can be permuted:

**Definition 24** (p-category) A set of terms  $A$  is a *p-category* (w.r.t.  $\Delta$ ) if for any permutation  $\pi$  on  $A$  and  $\Delta$ -model  $M$ ,  $\pi(M)$  is a  $\Delta$ -model.

Note that two terms  $a$  and  $b$  are interchangeable if and only if  $\{a, b\}$  is a p-category. Note also that in all p-categories, all terms are interchangeable. However, the reverse is not true: not in all sets of constraints, all sets of interchangeable terms or all categories are



p-categories. As an example where the reverse does not hold, take *FOL* and add to it the following constraints:

1.  $\{I(a_i)\}_i$  is a converging sequence of real numbers.
2.  $\{I(b_i)\}_i$  is a non-converging sequence of real numbers.

Call the new set of constraints  $\Delta$ . Now let  $A = \{I(a_i)\}_i \cup \{I(b_i)\}_i$ . Note that every two terms in  $A$  are interchangeable: a finite number of switchings between terms will not make a difference to the convergence of sequences. However,  $A$  is not a p-category. Let  $\pi$  be a permutation on  $A$  that switches between  $a_i$  and  $b_i$  for all  $i$  (“all at once”, so to speak). Let  $M$  be some  $\Delta$ -model. Clearly,  $\pi(M)$  is not a  $\Delta$ -model.

Looking at the example we can note a special feature of  $\Delta$ : there are constraints that involve infinitely many terms. Indeed, in any system of constraints where only finitely many terms are constrained by each constraint, all sets of interchangeable terms are p-categories. We prove this claim below. First, however, we formulate more precisely the requirement that each constraint constrains only finitely many terms. Recall that the notion of constraining a phrase was defined in Section 7.3.1 (Definition 5 on p. 130). Note that all constraints mentioned in this chapter satisfy this requirement. Thus, in the systems of constraints mentioned, categories are maximal p-categories.

(FIN) *Each semantic constraint constrains at most finitely many terms with respect to  $\emptyset$  (cf. Definition 5 on p. 130).<sup>19</sup> Moreover, for each constraint  $C$  there is a finite set of terms  $A$  such that for any set of terms  $B$  that  $C$  constrains with respect to  $\emptyset$ ,  $B \cap A \neq \emptyset$ .*

**Proposition 8** *Let  $A$  be a set of terms of a language  $L$  and  $\Delta$  a system of semantic*

---

<sup>19</sup>That is, there are finitely many terms  $t$  such that  $C$  constrains  $\{t\}$ . The number of *phrases* constrained might be infinite, and will usually be if each term can appear in infinitely many phrases.

## CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK

*constraints satisfying FIN.  $A$  is a  $p$ -category w.r.t.  $\Delta$  iff every two terms in  $A$  are interchangeable in  $\Delta$ .*

*Proof.* The left to right direction follows immediately from the definitions.

Right to left:

We shall use the following two claims about permutations:

(1) Let  $\pi$  be a permutation on a finite set  $A$ . There is an  $m \in \mathbb{N}$  and a series of permutations on  $A$   $\pi_1, \dots, \pi_m$  such that each  $\pi_i$  is a “switching” – a permutation that merely switches two terms in  $A$  and is constant on all other terms, and such that for each  $a \in A$   $\pi_m(\dots(\pi_1(a))\dots) = \pi(a)$ . (Informally: every permutation on a finite set is the result of the application of finitely many switchings.)

(2) Let  $\pi$  be a permutation on a set  $A$ . Let  $A'$  be a finite subset of  $A$ . Then there is a permutation  $\bar{\pi}$  on  $A'$  such that for each  $a \in A'$ , if  $\pi(a) \in A'$  then  $\bar{\pi}(a) = \pi(a)$ . (Informally: for any permutation on a set  $A$ , and for any finite subset  $A'$  of  $A$ , there is a permutation on  $A'$  that preserves the original permutation on the terms that don't get sent out of  $A'$ .)

I leave the proof of these claims to the reader.

Now let  $M$  be a  $\Delta$ -model. Let  $\pi$  be a permutation over  $A$ . We show that  $\pi(M)$  is a  $\Delta$ -model. Let  $C$  be some constraint in  $\Delta$ . We show that  $\pi(M)$  complies with  $C$  (i.e.  $\pi(M)$

is a  $\{C\}$ -model). Since  $C$  is arbitrary, it will then follow that  $\pi(M)$  is a  $\Delta$ -model.<sup>20</sup>

By assumption FIN, there is a finite set  $A' = \{a_1, \dots, a_n\}$  such that for any set of terms  $B$ , if  $C$  constrains  $B$ , then  $B \cap A' \neq \emptyset$ .

Let us extend  $\pi$  to be a permutation on all terms in  $L$ , so that it is the identity on all terms not in  $A$ .

Let  $A'' = \{\pi(a_i) : 1 \leq i \leq n\}$ . By claim (2) there is a permutation  $\bar{\pi}$  on  $A' \cup A''$  such that  $\bar{\pi}(a_i) = \pi(a_i)$  for all  $1 \leq i \leq n$ .

By claim (1), there is an  $m$  and a series of switchings for  $A' \cup A''$   $\pi_1, \dots, \pi_m$  such that  $\bar{\pi} = \pi_m \cdots \pi_1$ . Since  $\bar{\pi}$  is the identity on terms outside  $A$ , we can assume that the  $\pi_i$ s only switch terms in  $A$ .

By our assumption, every two terms in  $A$  are interchangeable w.r.t.  $\Delta$ . So since  $M$  is a  $\Delta$ -model,  $\pi_1(M)$  is also a  $\Delta$ -model, and so on for the rest of the permutations. By direct induction,  $\bar{\pi}(M)$  is a  $\Delta$ -model. So  $\bar{\pi}(M)$  complies with  $C$ .

Now let  $B$  be the set of terms on which  $\pi(M)$  and  $\bar{\pi}(M)$  disagree, that is:  $B = \{b \in Terms : I^\pi(b) \neq I^{\bar{\pi}}(b)\}$ , where  $Terms$  is the set of terms in  $L$ .

By the construction of  $\bar{\pi}$ ,  $\pi(M)$  and  $\bar{\pi}(M)$  agree on all terms in  $A'$ , so  $B \cap A' = \emptyset$ .

So by assumption FIN,  $C$  does not constrain  $B$ . That is, for any model  $N = \langle D, I \rangle$ , if there is a  $\{C\}$ -model  $N' = \langle D, I' \rangle$  that agrees with  $N$  on all the terms outside  $B$ , then  $N$

---

<sup>20</sup> We rely here on the following rule: if  $M$  complies with each constraint in  $\Delta$ , then  $M$  complies with  $\Delta$ . This might seem to be an obviously valid rule. However, given the extent of freedom we would like to provide with respect to the metalanguage, this rule should be made explicit. It would rule out the following pair of constraints:

- If there is more than one constraint,  $I(a) \neq \emptyset$ .
- $I(a) = \emptyset$

Each of these constraints, taken alone, allows any model where  $I(a) = \emptyset$ , but together they rule out all such models (and all other models). Note that the problem is not the danger of paradox. What is common to cases that defy the said rule is that taking constraints in conjunction changes the conditions of their satisfaction by a model.

*CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK*

is a  $\{C\}$ -model. But  $\bar{\pi}(M)$  is a  $\{C\}$ -model that agrees with  $\pi(M)$  on all the terms outside  $B$ , therefore  $\pi(M)$  is a  $\{C\}$ -model.

□

*CHAPTER 7. SEMANTIC CONSTRAINTS: THE FRAMEWORK*

## Chapter 8

# Determinacy, Invariance and Fixing a Term

In this chapter I explore the relation between three different notions: determinacy, invariance, and the notion of fixing a term. I first discuss what it is to fix a term in the semantics of a formal system, that is, what is a logical term in the “shallow” sense. Much of the literature on logical terms is devoted to finding out which terms (or notions) should be fixed – but little is said about what fixing a term amounts to. I show, under some natural assumptions about what fixing amounts to, that some terms cannot be fixed in ordinary model-theoretic semantics in a manner faithful to their meaning. The meaning of these terms requires a more elaborate structure than ordinary models provide. I claim that this fact can be used as a reason for not accepting these terms as logical in the “deep” sense. I then propose that the less structure a term requires in order to be fixed, the more logical it is.

Next, I relate the idea of fixing a term to that of *determinacy*. The notion of determinacy was defined in the previous chapter as a relation between terms, and was used to define the notion of *logical term* in the “shallow” sense. In the previous chapter I suggested that

a term is to be considered to be fixed in the system if it is determined by  $\{\mathcal{D}\}$ , where  $\mathcal{D}$  denotes the domain in each model. Here I propose a more general definition of determinacy, so that any feature of a model can be seen as determining some terms. I then suggest that the less structure it takes to *determine* a term, the more logical it is.

Lastly, I bring the notion of *invariance* into the picture. Invariance criteria for logical terms were discussed in Part I. Here I relate the notion of invariance of an operation to the general notion of determinacy and to that of fixing a term. I rely on some virtues of invariance criteria, although I do not promote any specific one of them. Invariance criteria are a tool in determining how much structure a term needs in order to be fixed. Thus, rather than settling on one as a criterion for logicity, I use invariance conditions as a *measure* for logicity. The stricter the invariance criterion, the more logical the terms it allows. And to conclude, I show how the use of invariance criteria can be generalized to deal with semantic constraints that do not fix terms completely.

## 8.1 Fixing a Term

In Chapter 2 I have distinguished between logical terms in the “deep” sense and in the “shallow” sense. Logical terms in the shallow sense are relative to a system, and are those terms that are held fixed in the system. The deep sense refers to the nature of logicity, and presupposes that there is a principled distinction between logical and nonlogical terms. A term is logical in the deep sense if it has the distinguished feature of logicity, regardless of whether it is fixed in a system. Those who doubt the existence of a principled distinction between logical and nonlogical terms will object to the deep sense.

Presumably, in a correct system for logic, the logical terms in the shallow sense (those that are held fixed in the system) should be logical in the deep sense (assuming that we accept there is a principled distinction between logical and nonlogical terms). Recall that the contemporary discussion on logical terms arises from the lacuna in Tarski’s definition

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

of logical consequence: the definition draws on a distinction between logical and nonlogical terms, which is not given in the definition. But assuming that we have such a distinction at hand, Tarski's definition tells us to keep the interpretations of the logical terms of an argument fixed, and vary the interpretations of nonlogical terms, in order to tell whether the argument is valid (Tarski, 1936). In the setting in which Tarski defines logical consequence (that is, in (Tarski, 1936)), only one fixed domain is dealt with, i.e. all models have the same domain. Keeping a term fixed means that it gets the same interpretation—has the same extension—in all models. Moving to a model-theoretic framework with multiple domains, we lose this feature of the fixed terms. The usual way of fixing terms, as we shall see shortly, does not in general amount letting them have the same extension in all models.

Yet what it is to fix a term in a formal system is usually assumed, and on the rare occasion that it is explained, it is not given precise definition. The reason might be that distinguishing, in a given system, between logical and nonlogical terms *in the shallow sense*—terms which are fixed and those which are not—does not seem to pose any problem. In the systems dealt with in contemporary mathematical logic and in the philosophical discussions on logic, the fixed, logical terms have distinguished symbols, while the non fixed terms are assigned generic letters of some type, such as upper case English letters for predicates and relation symbols. There is just no reason to give particular symbols for the nonlogical terms of a given syntactic type, since they all behave just the same in the system.

Sher describes the difference in the ways logical and nonlogical terms are treated in semantic systems:

The non-logical terms are strongly variable: any formally possible denotation in accordance with their formal “skeleton” (i.e., their being individual terms, n-place relation of individuals, etc.) is represented in some model. The logical terms, on the other hand, denote fixed formal properties of objects, and their denotations are subject to the laws governing these properties. These terms



## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

are fixed not in the sense that they denote the same entity in each model (the denotation of the universal quantifier in a model with 10 elements differs from its denotation in a model with 11 elements). Rather, while non-logical terms are defined within models, logical terms are defined by fixed functions *over* models<sup>1</sup>. (Sher, 1996, p. 675)

Further, Sher adds:

Looking at the system as a whole, we may say that the primitive non-logical terms and their denotations in models constitute the *base* of the system; the logical terms and their semantic definitions - its *superstructure*. The two parts of the system are brought together by “superimposing” the logical apparatus on the non-logical base. Syntactically, this is done by the definition of well-formed formulae (wffs), in which the logical terms are distinct formula building operators; semantically, by the definition of “truth in a model” which is based on (i) the logical structure of wffs, (ii) the model-theoretic definitions of the logical-terms. (Sher, 1996, pp. 675f)<sup>2</sup>

Sher does not make the distinction between the deep and the shallow sense of logical terms, but clearly, what she is describing here is the latter.<sup>3</sup> Although Sher captures, I think,

---

<sup>1</sup>Sher does not merely mean that the extension of a logical term is a function of models—such is the also the extension of a nonlogical term. The idea, I presume, is that the interpretation of a nonlogical term is a function of the domain.

<sup>2</sup>See also (Sher, 1991, p. 46f).

<sup>3</sup>One can also look at this quote as describing some principled distinction between logical and nonlogical terms, as is done in (Varzi, 2002). Varzi claims that the characterization of logical terms as interpreted outside the system of models and of nonlogical terms as interpreted within cannot be the basis for a principled distinction between logical and nonlogical terms, since any term can be either interpreted from either outside or within. But then, in the above quote, Sher does not propose a criterion for logicity: she only describes how logical terms are treated in Tarskian systems. Later in the same paper (and likewise in the book, (Sher, 1991)), Sher provides a deep analysis of logical terms as formal and general, and based on that she proposes

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

what is meant by the shallow sense of logical terms, she does not offer a precise, formal definition. Since the shallow sense already poses logical terms in a formal setting—only there it is intelligible—it is only fitting that there would be a formal definition for what it is to be fixed. But such a definition is not given in discussions of logical terms, perhaps because the idea seems clear and obvious enough as described by Sher. Clearly, however, understanding the shallow sense of logical terms, namely, what it means to be fixed in a semantic system, is pertinent to such discussions.

Now it would seem that the *completely* fixed terms of a language in a given systems are simply those whose extension is fixed: constant across all models (hence they are logical *constants*). This notion of fixing goes well with the precursors to standard contemporary model-theoretic semantics: *substitutional semantics* and single domain model-theoretic semantics—but not with standard, multiple domains model-theoretic semantics. Let us examine these three settings in turn.

For simplicity, let us focus on logical truth rather than logical consequence. In *substitutional semantics*, the logical truth of a sentence is determined by the truth of a range of sentences that share the same form, that are substitution instances of the original sentence. A substitution instance of a sentence is received from the original one by substituting nonlogical terms for nonlogical terms of the same syntactic category in a uniform manner. In substitutional semantics, a sentence is logically true if and only if all its substitution instances are true. There is no use of models here, just the material truth of a range of sentences. The logical terms are fixed here in the sense that they are not substituted.

Single domain model theory is simply ordinary model theory where there is just one domain, the universe of all existing objects. In the case of single domain model theory, we have a Tarskian reduction of logical truth to material truth. All nonlogical terms in the sentence are substituted by variables of appropriate type in a uniform manner, and universal quantifiers for all the new variables are appended to the sentence. The original

---

her criterion which tells *which* terms should be fixed from outside the system of models.

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

sentence is said to be a logical truth if and only if the resulting sentence is materially true (Tarski, 1936, pp. 416f). Here the logical terms are again those that are not substituted. Furthermore, if we consider the logical terms in single domain semantics to have extensions, as we did in previous chapters, we can say that logical terms in single domain semantics are *completely* fixed in the sense that they have the same extension in all models.

Now in the case of ordinary multiple domain model theory, as Sher indicates in the first quote, logical terms do not seem to be *completely* fixed. That is, for a logical term  $a$ , we don't have  $I(a) = I'(a)$  for all interpretations  $I$  and  $I'$ . With accepted definitions, this gives us the truth-functional connectives, but would leave out identity and the quantifiers. In Chapter 7 I presented logical terms in a way that follows accepted definitions (in a system of semantic constraints). The connectives were construed as functions from truth values to truth values, the same functions in each model. The quantifiers, on the other hand, were construed as second-level predicates, their extension not identical in all models. The extension of each of the standard quantifiers, ' $\forall$ ' and ' $\exists$ ', is different in models that have different domains. For any models  $M = \langle D, I \rangle$  and  $M' = \langle D', I' \rangle$ ,  $I(\forall) = \{D\}$  and  $I'(\forall) = \{D'\}$ , so if  $D \neq D'$ , then  $I(\forall) \neq I'(\forall)$ . Likewise, if  $D \neq D'$  then  $I(\exists) \neq I'(\exists)$ .

Enderton considers the universal quantifier not as a logical term, but as a *parameter* (Enderton, 1972, p. 68). Steven Kuhn infers from the case of quantifiers that "Constancy is a matter of degree" (Kuhn, 1981, p. 491). Kuhn has a distinction analogous to the one I made between completely fixed and non-fixed terms: schematic terms are terms for which the class of interpretations allows to be assigned any object consistent with their type, constants are those terms for which the interpretation is the same in all models. According to Kuhn:

In classical logic, predicates and individual constants are purely schematic, truth-functional connectives are pure constants, and quantifiers lie somewhere in between (*ibid*).

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

Although I agree with Kuhn's claim, I think he misses an important point about connectives and quantifiers and what distinguishes them from predicates in standard logic. I will argue below that both connectives and quantifiers have their *meanings* fixed, even if not their extensions.<sup>4</sup>

Now, as far as Kuhn's definitions go, his description of standard logical systems is correct: not all terms are either purely schematic or pure constants. It is tempting to translate this observation into a claim about logical terms, at least in the shallow sense: that there is no strict dichotomy between logical and nonlogical terms in the shallow sense. But this claim is problematic, at least if there *is* a strict dichotomy between logical and nonlogical terms in the *deep* sense.

Assume that there is a principled, "deep" distinction between logical and nonlogical terms. Recall that in a proper logical system, logical terms in the shallow sense are supposed to track the deep distinction: a logical term in the shallow sense should be logical in the deep sense. But what Kuhn gives us is a spectrum of logicity in the shallow sense. How does a graded notion track a non-graded notion? To what degree do we fix the logical terms in the deep sense? A line has to be drawn somewhere, on one side of which will lie the logical terms and on the other the nonlogical terms. Both connectives and quantifiers are considered as logical in the deep sense on most accounts, but according to Kuhn, they are fixed to different degrees. Kuhn, however, leaves us with a missing parameter telling us to what extent a logical term in the deep sense should be fixed. Criteria for logical terms don't provide such a parameter, as they tend to give stringent conditions and aim at an exclusive collection of terms. They don't give grounds for making distinctions among logical terms and for fixing some more than others. Kuhn does not supply such grounds either.

---

<sup>4</sup>The idea that there are degrees of constancy is consonant with the idea behind the framework of semantic constraints. I shall relate this discussion to semantic constraints only at the end of this chapter. I will just note here that the framework I am proposing takes a step further, allowing *meanings*, rather than extensions to be fixed to various degrees.

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

Perhaps Kuhn is not committed to the idea of logical terms in the deep sense. But I think that Kuhn simply misdescribes the terrain vis-à-vis logical terms in standard systems, which is compatible with the distinction between logical and nonlogical terms in the deep sense. There is a sense in which quantifiers *are* fixed, and is not simply their being non-purely-schematic. Given a domain, there is only one possible extension for each of the quantifiers. This property is common to all logical terms in standard extensional systems. Thus, it is not the *extension* that is held constant when we fix a term, but rather its *intension*, or its *meaning*. ‘ $\forall$ ’ does not have the same extension in all models, but it always means “all objects” or “all of the domain”. Likewise with ‘ $\exists$ ’ and ‘ $=$ ’. As I will show later on, there is a difference between connectives and quantifiers with respect to *what it takes* to fix their meaning, that is, how much structure they require, and there we do find varying degrees.

Talk of intensions is not out of place in this setting, even though it is standard extensional logic we are concerned with. If our formal system is to bear any important relation to correct reasoning, the terms of the formal language should be associated with some meanings—at least those terms that are fixed. Let an intension be a function from possible worlds to extensions. In order to accommodate intensions in this setting, we shall look at the models (still in extensional logic) as representing possible worlds, as was done in Chapter 3. Recall that in Chapter 3 we analyzed a modal element in logical consequence, that led us to the approach that models represent possible worlds in some way or another. This assures us that we capture the condition of *Necessity* (formulated on p. 20).<sup>5</sup> By associating models with possible worlds we make sure that we capture the modal element in logical consequence.

We shall make use of the idea that models represent possible worlds to accommodate

---

<sup>5</sup>In this work I the notion of *metaphysical necessity* is assumed rather than analyzed. I assume that this notion can be dealt with through the use of *possible worlds*, but other options, such as *possible situations* are not in principle excluded.

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

intensions. To avoid confusion, I shall refer to the collection of objects in a possible world as its *population*, to be distinguished from domains of models. It then makes sense to look at domains of models as representing populations of possible worlds (even though this might not be a one-one correspondence). The function from domains to extensions associated with a fixed term thus represents an intension. This intension can be associated with the term beforehand informally, as is usually the case—and then there is a matter whether the model theory captures this intension adequately. If the term is not supposed to capture a preconceived meaning, we can simply define a new term through the way that it is fixed in the model theory.

Presumably, debates on logical terms are about which terms should have their meaning fixed across models. Proponents of the demarcational project contend that only terms that are logical in the deep sense should be fixed in this way. This would make it seem that for any term with an associated intension, we could either fix it or not, and we should fix it only if the fixed term satisfies the proposed criterion for logical terms. But the semantical framework we are using, of standard extensional logic, imposes considerable restrictions on what *can* be fixed. First, there are the standard examples of intensional operators, such as *necessarily*, *ought* and *believed*. Their meaning is not a function of the population of the possible world or situation in which they are evaluated: that is why each model in semantic systems for these operators includes an elaborate structure of constructs called “possible worlds”.

Leaving the intensional operators aside (I shall discuss them in the next chapter), we are still left with a significant class of terms that cannot be fixed in the given framework. In general, any predicate that by the meaning we attach to it does not hold essentially of some object in some possible world cannot be fixed. The reason, as we shall see, is that for such predicates, you need more than the population of the world to fix its extension.

Let *Red* be a predicate in our language with the usual (or perhaps sharpened) meaning. I make the simplified, but I think harmless, metaphysical assumption that it is possible for

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

any object extended in space to be red: for each object  $o$  there is a possible world  $u$  in which it is red, and a possible world  $u'$  in which it is not red. I assume further, for simplicity, that the redness of an object is independent of what other objects exist.

Now *Red*, on most accounts of logicity, is regarded as a paradigmatic example of a nonlogical term (in the deep sense). In a semantic system where *Red* is *not* fixed, we would have various models, accounting (also) for the variety of extensions *Red* can obtain. As we have seen in Chapter 3, the range of models can be seen to represent both variations in possible worlds and variations in the meanings of the nonlogical terminology, and together with Shapiro, we were skeptical as to whether a line can be drawn between the two types of variation.

Now if we were to consider fixing *Red* as a logical term in the shallow sense, we would run into trouble. Let us take a simplified example. Let  $D = \{1, 2, 3\}$  be a domain representing the population of some world or some worlds with just three objects:  $a$ ,  $b$  and  $c$ , so that 1 represents  $a$ , 2 represents  $b$  and 3 represents  $c$ . Assume that at least  $a$  is concrete and is extended in space. If we want to fix *Red*, we should assign it to members of the domain according to its intended meaning, that is, according to whether the objects they represent are red. But the domain and the objects it represents do not suffice to determine the extension of *Red*. According to our assumptions, there is a world  $w$  whose objects are  $a$ ,  $b$  and  $c$ , where  $a$  is red, and a world  $w'$  with the same objects, where  $a$  is not red. There is no way of telling from just  $D$  and the population it represents whether or not to include 1 in the extension of *Red*. The way we set things up so far, *Red* is simply *precluded* from being a logical term in the shallow sense.

The same problem does not arise for standard logical terms. The intension of the universal quantifier, understood as “all objects”, will give the same extension in possible worlds consisting of the same objects, regardless of their properties and relations. Take the function  $f_{\forall}$  taking each domain  $D$  to the extension of ‘ $\forall$ ’ in that domain,  $\{D\}$ . This function is not a one-one representation of the intension associated with ‘ $\forall$ ’—but it is a

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

good enough representation, since it doesn't lose any of the intended meaning of ' $\forall$ ' — it does not erase any of the distinctions ' $\forall$ ', on its intended meaning, could make. Let us say that a term is *fixed faithfully* if it is fixed, namely its extension is a function of the domain, and its intension is constant on all possible worlds that happen to be represented by one domain in the model theory.

As an aside, note that we run into less trouble with a term denoting an essential property. Assume that being human is an essential property: an object has it in some world only if it has it in all worlds. Then the extension in a model of the term *Human* with its intended meaning is just a function of which objects are represented by the domain of the model. If the objects represented by a domain are a function of the domain, *Human* can be fixed faithfully in the present framework. There would be some trouble, however, if we have a one-many relation between domains of models and populations of possible worlds they represent, in case some element of a domain represents a human in one world and a non-human in another. Thus, fixing a term denoting an essential property can be done faithfully, although it could, in some cases, set limits on the relation between models and possible worlds. As for other grammatical categories, if there are essential functions, relations or objects, those could be fixed faithfully as well. "Father of" might be an essential function and numbers might be essential objects (i.e. objects that necessarily exist), and in that case we can fix the terms denoting them. However, pursuing this idea will take us too far afield.

An aspect of the same issue is brought up by Menzel (1990). Menzel discusses modal logic, and the ability of clusters of models to characterize modal reality. That is, whether we can take a bunch of standard models of extensional logic to represent possible worlds. Menzel's discussion is highly relevant to the issue at hand.

In order for our model-theoretic notion of consequence to be adequate, models should capture modal reality in some way. "[N]ote", Menzel tells us, "that things could have been different in one of two ways: there could have been more or fewer individuals, or the



## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

individuals there are could have had different properties and stood in different relations to one another. Distinct plain vanilla models can capture the first sort of difference well enough (simply in virtue of having different domains), but not the second” (Menzel, 1990, p. 359). Going back to the example in the previous page:  $D$  represents only populations of possible worlds with three objects, indeed, only those with the objects  $a$ ,  $b$  and  $c$ ; but  $D$  cannot distinguish between  $w$  and  $w'$  where the objects have different properties. Now when the logic-as-model view was presented, it was stressed that the representation of possible worlds does not have to amount to a full description, *as long as we get the relation of logical consequence right*. Thus, we don’t always have to represent the second sort of difference Menzel mentions. In the usual case, where *Red* is not fixed, there is no need to model the way *Red*’s intended meaning should behave in each model. Using Kuhn’s terminology, *Red* in that case is purely schematic – it then receives any possible extension that accords its grammatical category in some model. The interpretation function can then be seen to account for variations in the properties and relations between objects in the world as well as of variations in the interpretation of nonlogical vocabulary, and there is no reason to distinguish between the two (see Chapter 3). But here the situation is different: we want to fix *Red* as a logical term, that is, account for all metaphysical possibilities with respect to being red, while keeping the meaning of *Red* fixed. So here we can’t let the interpretation function blend variations in possible worlds with variations in meaning. We need to somehow extract the meaning of *Red* and keep it fixed, while still accounting for all possibilities that have to do with it.

It might be claimed that the modality is meshed with meaning and there is no point in trying to separate them. Or at least so when it comes to some of the terms in the language: otherwise no term could be fixed. So some terms, say those that have to do with contingent properties or relations, cannot be fixed in the semantics under this view of things. We would then get an *ex ante* condition for a term to be logical coming from the limitations of the semantics. I will not claim one way or another regarding the relation between meaning and

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

modality. But if those two *can* be separated, then there are two salient formal solutions to the problem of fixing *Red* and similar terms, which are presented below. In both cases we limit our concern to the fixity of meaning of predicates and relation symbols.

The first solution to the problem of fixing the meaning of terms is presented by Menzel. The solution is to add more structure to the models, so that even before the interpretation function comes in, models with the same domains can still be distinguished. The idea, modified to the current setting, is that each model will be a triple  $\langle D, P, I \rangle$  such that  $D$  and  $I$  are the usual domain and interpretation function, and the added element  $P$  is function from terms to extensions. The first two elements in a model represent a possible world, including a population of objects in that world, represented by  $D$ , and the properties and relations between the objects, represented by  $P$ .  $P$ , just like the interpretation function  $I$ , assigns an extension to each term  $t$ . But unlike  $I$ ,  $P$  makes no difference between logical and nonlogical terms. For each term  $t$ ,  $P(t)$  is the extension of  $t$  under  $t$ 's intended meaning. That is,  $P(t)$  is the extension of  $t$  in the possible world represented by the model, given the meaning  $t$  bears. The interpretation function  $I$  also assigns an extension to  $t$ . If  $t$  is fixed in the semantics,  $I$  should assign  $P(t)$  to  $t$ , and likewise in all other models (cf. (Menzel, 1990, pp. 360f)<sup>6</sup>). Thus, on the solution we take from Menzel (with some superficial modifications), the original interpretation function is split into two: a function

---

<sup>6</sup>I have made some changes to Menzel's proposal to fit it into the current discussion. Menzel's models (for an extensional language) are triples  $\langle D, P, Ext \rangle$ , where  $D$  is the domain,  $P = \{P_n : n \in \mathbb{N}, n > 0\}$  is a set of pairwise disjoint sets (where each  $P_n$  stands for the set of  $n$ -place relations), and  $Ext$  a function assigning a subset of  $D^n$  to every member of  $P_n$ . A valuation function for a language  $L$  assigns a member of  $D$  to each singular term of  $L$ , and a member of  $P_n$  to each  $n$ -place predicate. In an arbitrary model, the valuation function can assign any member of  $p \in P_n$  to a given  $n$ -place predicate, but in the *intended* model,  $p$  and  $Ext(p)$  will match the intended meaning of the predicate (Menzel, 1990, pp. 360f). Thus, only intended models with respect to some predicate can be said to fix its meaning. The models in the revised version of Menzel's proposal, presented in the main text, are all intended—the others are not included—as we are interested in fixing the meaning of terms across all models.

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

that assigns extensions to terms according to their meaning given the possible world that is represented, and a function that assigns extensions to terms – any extension that fits the syntactic category if the term is not fixed, and the one extension that accords with the intended meaning if the term is fixed.<sup>7</sup>

Menzel’s solution requires modifying the model theory: standard models on their own do not suffice to determine a possible world, and therefore another function is added as an element in models. But perhaps we can bypass this revision to model theory, and still be able to fix a term such as *Red*. Let us move on to the second proposed approach to fixing terms, where models remain intact. Note that we have assumed so far that the domain of a model represents the population of some possible world. That is, each element in the domain of a model represents an object in the relevant possible world. Maybe we can lay some more burden on the relation between elements of domains and the objects they represent. When we have two possible worlds with the same populations that only differ in the properties or relations among their objects, we will use different domains to represent them. An element of a domain will not simply represent an object, but rather an object bearing certain properties and relations to other objects.

The idea is to code the properties and relations into the elements of the domain. Let us use ordinals to represent objects in possible worlds (assume that we have enough ordinals). Each ordinal that is used will code both which object is represented by it and the properties it has and relations the object bears to other objects. Thus, each object that appears in different possible worlds will be represented by different ordinals, according to the properties

---

<sup>7</sup>A similar (though much more general) solution to the problem of fixing arbitrary terms, of adding more structure to models, is presented by Varzi (2002). Varzi claims that “one could treat *any* given term outside the system of models or inside it, as the case may be, provided that a semantic apparatus is available that is sufficiently general and unbiased to support such practices” (Varzi, 2002, p. 204, my emphasis). That is, Varzi contends that any term can be fixed. However, Varzi does not consider a modal condition on logical consequence (such as *Necessity*), and does not show that his proposal abides by any such condition.

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

it has and the relations it bears to other objects in each of the possible worlds. I will not give the function that does this: this of course cannot be done without a systematic presentation of objects of possible worlds, which we lack. Similar to the situation before, where we assumed that each model represented a possible world without telling exactly how it is done, we assume here that more complicated relation of representation is also feasible.

To illustrate this idea, let us assume that we have a function that codes, with ordinals, objects and their properties and relations. The function has two arguments: the first is an object, and the second a list of the properties it has and relations it bears to each other object. The function is injective, so that different objects or an object with different properties or relations will be coded by different ordinals. It will suffice to consider in this list only properties and relations that are denoted by terms in the language we wish to fix. Nevertheless, this list will in general be infinite, as it will include an item for any other object to which the object bears a relation. Now let  $a$  be some object, say an apple, in some possible world  $w$ . Assume that in  $w$ ,  $a$  is red, tasty, and on the table  $b$ . We can write down the ordinal that represents it as  $\alpha_{Red(-),Tasty(-),On(-,b)}^a$ . The superscript of  $\alpha$  represents the object, and the subscript represents the properties and relations it bears. This is a simplified example, as I have considered only two properties and one relation. Let us consider a more abstract and slightly more complicated example to show how different domains represent the same population in different possible worlds. Say that we have the possible world  $w_1$  with objects  $a$  and  $b$ , such that the predicate  $t_1$  under its intended meaning holds of  $a$  but not  $b$ , and the binary predicate  $t_2$ , under its intended meaning, holds between  $a$  and  $a$  and between  $a$  and  $b$ . Assume that  $w_2$  has the same objects as  $w_1$ , but there  $t_1$  holds of  $b$  but not  $a$ , and  $t_2$  holds just between  $a$  and  $b$ . We can use the following domains to represent these possible worlds:  $D_1 = \{\alpha_{t_1(-),t_2(-,-),t_2(-,b)}^a, \alpha_{t_2(a,-)}^b\}$  and  $D_2 = \{\alpha_{t_2(-,b)}^a, \alpha_{t_1(-),t_2(a,-)}^b\}$ . In  $D_1$ , for instance,  $a$  is represented by  $\alpha_{t_1(-),t_2(-,-),t_2(-,b)}^a$ , where the indexing indicates the properties and relations  $a$  has in  $w_1$ . In  $D_2$ ,  $a$  is represented by  $\alpha_{t_2(-,b)}^a$ , and the difference in index indicates that  $\alpha_{t_1(-),t_2(-,-),t_2(-,b)}^a$  and  $\alpha_{t_2(-,b)}^a$  are different ordinals. Now if we want to fix the

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

meaning of some term, say  $t_1$ , we impose that in each model  $\langle D, I \rangle$ ,  $I(t_1)$  has as members only the elements of the domain that represent objects having  $t_1$  on its intended meaning in the possible world depicted by  $D$ . So if we fix  $t_1$ , in any model with  $D_1$  as domain, the interpretation of  $t_1$  will be  $\{\alpha_{t_1(-), t_2(-, -), t_2(-, b)}^a\}$ , and in any model with  $D_2$  as domain, the interpretation of  $t_1$  will be  $\{\alpha_{t_1(-), t_2(a, -)}^b\}$ . The examples just given are very simplified, since, as earlier mentioned, the subscripts will in general consist of infinite lists. Those lists will include non-binary relations that were not considered here. Yet an injunctive function into the ordinals is still possible in principle, through which we can fix terms according to their intended meaning.

The second solution to the problem of fixing terms keeps models in their standard form, on the expense of complicating the way models represent possible worlds. In both solutions, the extension of a fixed term is a function of some elements of the model representing a possible world. In the Menzel solution, the extension of a fixed term is a function of  $\langle D, P \rangle$ . In the second solution we offer, the extension of a fixed term is a function of  $D$  alone. This way we can bring *Red* to the same field as  $\forall$  and  $\exists$ . Menzel's solution, although complicating the models, is simpler and carries greater heuristic value.

Now after we have shown how any term might be fixed, it becomes obvious that fixing a term like *Red* involves a fair amount of metaphysics. For instance, the logical status of  $\varphi := \neg \forall x Red(x)$ , which is neither logically true nor logically false when *Red* is not fixed, becomes unclear when *Red* is fixed (assuming  $\forall$  is fixed). Presumably,  $\varphi$  would not be logically false when *Red* is fixed: there is a possible world where not everything is red. However, it's more difficult to tell whether  $\varphi$  is logically *true*: is there a possible world where everything is red? Answering this involves more metaphysics than the present author is informed with. This might be a reason to rule out fixing *Red* as a logical term. We might extract here a new criterion for logical terms (in the deep sense), though at this point in a vague formulation: a term is logical if the machinery required for fixing it is metaphysically acceptable and easy to handle. This criterion is in line with the way we

motivated *formality* in Section 3.2, as the element in logical consequence that helps avoid metaphysical questions. I shall make it more precise in the following sections. Now I would rather point out that given some extra structure in our model theory, any term could in principle be fixed. Allowing extra structure, the obstacle to fixing some terms is our shaky handle on the metaphysics involved.

## 8.2 Fixity and Determinacy

Up to this point, using some examples, I have spoken rather loosely about the structure it takes to fix a term faithfully. In the remainder of the chapter, I employ the notions of *determinacy* and *invariance* to make the ideas more precise. I shall start by bringing determinacy into the picture (to be defined below). We shall see that a term that does not need much structure to be fixed faithfully is a term that is determined by very general features of models when fixed faithfully.

In Chapter 7, where the framework of semantic constraints was presented, I characterized logical terms in the shallow sense using the notion of determinacy relative to a system of semantic constraints. All the standard “term-based” systems—those that rely on a distinction between logical and nonlogical terms—can be construed in the framework of semantic constraints, so we will not lose any generality going back to that definition. A term  $a$  is determined by a set of terms  $B$  (relative to a set of constraints  $\Delta$ ) if every possible extension for the terms in  $B$  fixes one possible extension for  $a$ : the extension of  $a$  is a function of the extension of the terms in  $B$  (in the  $\Delta$ -models). We add a pseudo-term  $\mathcal{D}$  to any language with the constraint:  $I(\mathcal{D}) = D$ . In this way we can define the fixed terms, or the logical terms (in the shallow sense), as those that are determined by  $\{\mathcal{D}\}$ .

This definition is in accord with standard usage, at least in the frameworks we have been considering so far, of standard extensional logic. In what follows I generalize the notion of determinacy so that it would deal with various features of models rather than just the

interpretation of terms, and pseudo-terms will not be necessary anymore.

- Definition 25** (i) Let  $\Delta$  be a set of semantic constraints. Let  $T$  be a partition of the class of  $\Delta$ -models for  $L$ . A term  $a$  is *determined by the partition  $T$*  if for any two  $\Delta$ -models  $M = \langle D, I \rangle$  and  $M' = \langle D, I' \rangle$ , if  $M$  and  $M'$  belong to the same class in the partition, then  $I(a) = I'(a)$ .
- (ii) Let  $f$  be some function on models (*function* is here used in a wide sense, not restricted to sets).  $f$  induces an equivalence relation  $\sim_f$  between models:  $M \sim_f M'$  iff  $f(M) = f(M')$ .  $\sim_f$  induces a partition  $Models / \sim_f$  on the class of models. We say that  $a$  is *determined by the function  $f$*  if  $a$  is determined by the partition  $Models / \sim_f$ .

Examples:

1. First let us connect the more restricted notion of determinacy to the generalized one, and see how the former can be made a special case of the latter. Let  $b$  be some term in  $L$ . Let  $f_b$  be the function that gives to each model the extension of  $b$  in that model as value. Namely, for  $M = \langle D, I \rangle$ ,  $f_b(M) = I(b)$ . For any term  $a$ ,  $a$  is determined by  $\{b\}$  if and only if  $a$  is determined by  $f_b$ . To deal with determination by a non-singleton set of terms, let  $B$  be a nonempty set of terms in  $L$ . Let  $f_B = \{I(b) : b \in B\}$  for  $M = \langle D, I \rangle$ . Then  $a$  is determined by  $B$  if and only if it is determined by  $f_B$ .
2. As a special instance of the previous example, we can redefine “determination by the domain” in the new setting. Let  $f_D$  be the function that gives to each model its domain as value. Namely, for  $M = \langle D, I \rangle$ ,  $f_D(M) = D$ . For any term  $a$ ,  $a$  is determined by  $\mathcal{D}$  (the pseudo-term denoting the domain) if and only if  $a$  is determined by  $f_D$ ; for simplicity we say in such a case that  $a$  is *determined by the domain*.

When fixed with the usual syntactic and semantic clauses,  $\forall$ ,  $\exists$  and the truth-functional connectives are determined by the domain.

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

3. To complete the reduction of the restricted notion of determination to the generalized one, we turn to determination by the empty set (of terms). Let  $T$  be the partition of models with just one class, the class of all models. For any term  $a$ ,  $a$  is determined by  $T$  if and only if  $a$  is determined by the empty set.

When fixed with the usual syntactic and semantic clauses, the truth-functional connectives are determined by  $T$ . The quantifiers  $\forall$  and  $\exists$  are *not* determined by  $T$ . However, the quantifier  $\nexists$ , meaning *there is no* or *none*, is determined by the  $T$ , when fixed according to its intended meaning by the clause:

- $I(\nexists) = \{\emptyset\}$

4. For the last example we modify the model theory according to Menzel's proposal under our modifications. So each model is a triple  $\langle D, P, I \rangle$  of a domain, the intended interpretation function, and the ordinary interpretation function. In this setting, as we have seen, the meaning of arbitrary terms can be fixed in a way that is faithful to their intended meaning. Being fixed, in this setting, can be explicated as *being determined by*  $f_{\langle D, P \rangle}$ , in short: *being determined by*  $\langle D, P \rangle$ , where for  $M = \langle D, P, I \rangle$   $f_{\langle D, P \rangle}(M) = \langle D, P \rangle$ .

In the discussion on fixity in the previous section, we came to the following conclusion: In the standard setting, where models are pairs of domains and interpretation functions, and the domain of a model is seen as representing the population of some possible world, a term such as *Red* cannot be fixed according to its intended meaning. Implementing the notion of determinacy we may rephrase things as follows: in a semantic system in the standard setting as just described, if the term *Red* is determined by the domain, then *Red* is not faithful to its intended meaning. *Red* cannot be a logical term in the shallow sense in a system according to the definition of logical terms from the previous section. The situation changes when we add more structure to the models or the way they represent possible worlds.



## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

In the previous section I considered two solutions to the problem of fixing terms such as *Red*. On the first, taken from Menzel, we add more structure to the models. And then *Red*, when fixed according to its intended meaning, would be determined by  $\langle D, P \rangle$ , as explained in the last example. On the second solution, the model theory was left intact, but we found a way for a domain on its own to represent a possible world including both its population and the properties and relations that hold in it. In that setting, we can have *Red* fixed according to its intended meaning and be determined by the domain.

On both solutions, *Red* is determined by the aspect of a model that represents a possible world in full. This is what we should expect. If the intended meaning of a term, its intension, is a function from possible worlds to extensions, then the amount of structure that takes to represent a possible world in full should suffice to determine the extension of any fixed term. We may add, extending for a moment the scope of discussion, that possible worlds could be represented in an even fuller manner, accounting for accessibility relations between them and other worlds. A structure representing those aspects of possible worlds would be enough to determine the usual modal terms *necessarily* and *possibly*.

To conclude the discussion so far, we note that terms can be distinguished by the amount of structure that takes to fix them in a manner faithful to their intended meaning (if such exists). The terms standardly taken as logical terms need less structure than do other terms, for instance, terms denoting contingent properties whose intension function gives different results for two possible worlds that have the same objects.

Recall that the definition of logical terms through determinacy was intended to capture the *shallow* sense. But here a scale for the logicity of terms, in the *deep* sense, suggests itself. We can rephrase the idea from the end of the previous section in more precise terms: the less structure that is required to determine a term in accordance with its intended meaning, the more logical it is. This idea is given more substance in the next section, where fixity and determinacy are connected to criteria of invariance for logical terms. We shall see that the suggested criterion can be accommodated in the general framework of semantic

constraints, where we will obtain a scale for the logicity of semantic constraints rather than just terms.

### 8.3 Invariance and Determinacy

To end this chapter, I relate the notions of *fixity* and *determinacy* to invariance criteria for logical terms. Recall that invariance criteria were introduced as criteria for logical *notions* or *operations*, but also indirectly for logical *terms*. In the previous sections I made a connection between terms, their intended meaning (their intension) and the operations they denote. We shall here apply invariance criteria to terms, given an associated operation.

Invariance criteria provide us with a function or a relation between models, and terms that are invariant under those, pass the test for logicity. As we shall see, the same test also tells us that these terms can be determined by a certain general feature of models. In this section I observe how various invariance criteria can be reformulated through the notion of determinacy, and then serve as a basis for evaluating the degree logicity of terms.

In Part I we have seen a variety of invariance criteria for logical terms (in the *deep* sense). Take, for instance, the Tarski-Sher thesis, according to which logical terms are invariant under isomorphisms. That is, a logical term behaves the same in isomorphic models. Now, “behaves the same” does not mean that its extension is the same. A term  $t$  is invariant under isomorphisms if for any models  $M = \langle D, I \rangle$  and  $M' = \langle D', I' \rangle$ , if  $f$  is a bijection from  $D$  onto  $D'$ ,  $f(I(t)) = I'(t)$ . On first approximation we can say that the isomorphism “type” of a model suffices for telling us how the term behaves in the model.

From now on I assume the notational rule that where a model is signified by an ‘ $M$ ’ with some superscript, its domain and interpretation function are signified by ‘ $D$ ’ and ‘ $I$ ’ with the same superscript, so  $M = \langle D, I \rangle$ ,  $M' = \langle D', I' \rangle$ ,  $M^* = \langle D^*, I^* \rangle$  etc.

The relation  $\leftrightarrow$  between models, defined as:

$$M \leftrightarrow M' \text{ iff there is a bijection from } D \text{ onto } D'$$

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

is an equivalence relation. So  $M \leftrightarrow M'$  if and only if the domains of  $M$  and  $M'$  have the same size. The function  $card$ , defined by:

$$card(M) = \text{the cardinality of } D$$

defines a partition of the class of models, the same as the one derived from the relation  $\leftrightarrow$ , since  $M \leftrightarrow M'$  iff  $card(M) = card(M')$ .

Now the isomorphism-invariance criterion for logical terms says that logical terms are invariant, or interpreted in the same way, in models of the same equivalence class in the received partition. This invites using the notion of determinacy, and making a connection between invariance under isomorphism and determinacy by  $card$ . Indeed, in a sense, the cardinality of the domain is the structural feature that matters for isomorphism-invariant terms.

However, the connection is not straightforward. Invariance under isomorphisms is not equivalent to determinacy by  $card$ : the latter implies the former, but not vice versa. Take the quantifier  $\forall$ , interpreted by the clause:  $I(\forall) = \{D\}$ . There are models  $M$  and  $M'$  where  $card(M) = card(M')$  but  $D \neq D'$ , thus  $I(\forall) \neq I'(\forall)$ .  $\forall$  is not determined by  $card$ , though it is invariant under isomorphisms.

Let us consider now a slight revision to the model theory. We make the class of models thinner: in the new version, all models of the same cardinality have the same domain—so there is only one domain of each cardinality. In fact, since cardinal numbers are sets, we can simply use cardinal numbers as the domains of all models. For standard languages, this revision does not affect validity: the arguments that are valid in the ordinary version are exactly those that are valid in the thin one. This is a result of the fact that standard logic satisfies the *isomorphism property* (See Chapter 3 and (Shapiro, 1998)). So the thinner model theory is simply a more economic version of ordinary model theory, without what may be considered as an excess in models. In the thin version, however, isomorphism invariance *is* equivalent to determinacy-by- $card$ . The explanation is simple. Let  $t$  be an

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

isomorphism invariant term. So  $t$  is invariant under the identity bijection between domains: for any  $M, M'$  where  $D = D'$  we have  $I(t) = I'(t)$ . Now in our revised model theory,  $\text{card}(M) = \text{card}(M')$  implies  $D = D'$ . So  $\text{card}$  determines  $t$  in the thin model theory. That is,  $\text{card}$  serves as a general feature of models which determines the isomorphism-invariant terms. This result brings out clearly what it takes to fix isomorphism-invariant terms. With this result we capture the idea that the structural feature that takes to fix an isomorphism-invariant term is just the size of the domain—to be compared with a term such as *Red*, which, as we have seen, requires much more.

Considering the earlier discussion on what it takes to fix a term, we might worry whether the isomorphism-invariant terms can be fixed in the thin model theory in a manner faithful to their meaning—whether their meaning is not somehow violated when some models are lost. The claim before was that standard model theory is not capable of fixing *Red* faithfully, but the standard logical terms could nonetheless be fixed. The isomorphism invariant terms include the standard logical terms as well as others. The question now is whether those can be fixed faithfully in the thin, revised model theory. I claim that they can be faithfully fixed: precisely because they “behave” the same in all models of the same size in the standard system. Before, we had each domain in the standard model theory representing the population of some possible world or possible worlds. In the thin model theory, each domain can be seen to represent the populations of all possible worlds sharing its cardinality. And since cardinality is all that matters for isomorphism-invariant terms, the thin model theory suffices for fixing its meaning.

We come to the conclusion that, even though  $\text{card}$  does not determine isomorphism invariant terms in standard model theory, there is a revised model theory where it does. The revised version gives us all the structure we need to capture the full meaning of isomorphism invariant terms. We have learned that not only does the thin model theory give us the same validities as the standard one, it is faithful to the meanings of the logical vocabulary we standardly use. And this vocabulary is very economical in the amount of structure it requires

of the model theory—whereas other terms, such as *Red*, require much more structure to appropriately model their behavior in possible worlds. Indeed, Sher, as other proponents of invariance criteria, has claimed to have captured the *generality* of logical terms (Sher, 1996; Bonnay, 2008). We see here another aspect of this generality: isomorphism-invariant terms do not require commitment to a specific metaphysically elaborate set-up. They can be captured by a simple, mathematically and metaphysically undemanding model theory.

Stricter invariance conditions yield terms that require even less structure for being fixed adequately. Let  $\mathcal{F}$  be a class of functions, each one from some model to some model.  $\mathcal{F}$  can be the class of isomorphisms, or potential isomorphisms etc. Let  $\sim_{\mathcal{F}}$  be a relation between models such that  $M \sim_{\mathcal{F}} M'$  iff there is a function  $f$  in  $\mathcal{F}$  from  $M$  to  $M'$ . Let us assume that  $\mathcal{F}$  is such that  $\sim_{\mathcal{F}}$  is an equivalence relation between models (as is the case where  $\mathcal{F}$  is the class of all isomorphisms between models). We can thin out models, as before, keeping only one model from each  $\sim_{\mathcal{F}}$ -equivalence class. Let  $t$  be a term that is invariant under  $\mathcal{F}$ . As before, the model theory thinned by  $\sim_{\mathcal{F}}$  suffices to fix  $t$  in a manner faithful to its meaning: it has all the structure required to capture the meaning of  $t$ . In this thinned model theory, the partition  $Models / \sim_{\mathcal{F}}$  determines  $t$ : all sets in the partition are singletons. If  $\mathcal{F}$  includes all isomorphisms as well as other functions, we can say that  $\mathcal{F}$ -invariance requires *less* structure of the models than does isomorphism invariance.

Not all invariance criteria that have been offered give us a neat partition of the models as we obtain with the criterion of isomorphism-invariance. For example, the relation between models obtained by homomorphisms, Feferman's selected type of function (see Chapter 4.3 and (Feferman, 1999)), is not symmetric. However, there is a class of functions  $\mathcal{F}$  such that the  $\mathcal{F}$ -invariant terms are exactly the homomorphism-invariant terms, that does yield an equivalence relation between models (see (Bonnay, 2008, p. 42)). Similarly, for any class of functions that does not yield an equivalence relation in the straightforward way we used with isomorphisms, an adequate equivalence relation can nevertheless be obtained.<sup>8</sup>

---

<sup>8</sup>Let  $\mathcal{F}$  be a class of functions. Let  $Inv(\mathcal{F})$  be the class of all  $\mathcal{F}$ -invariant operations. And for any class

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

The more models we thin out, the representation relation between models and possible worlds becomes more abstract. It would be unnecessary and of negligible heuristic value to go through the process of thinning for the various invariance criteria. The basic idea remains: the invariance criteria proposed in the literature give us terms that need less structure than is offered by the standard model theory in order to be fixed. This is another way of stating their generality.

In Chapter 3 I described formality as a way to avoid metaphysical questions in logic. While it may be that we will always have to deal with *some* metaphysical questions, the objective is to keep those at a minimum, given other considerations. We have seen that the logic of isomorphism-invariant terms avoids a great many metaphysical questions, although it leaves the question of the sizes of possible worlds. We can thus connect invariance criteria in general to the view of formality as avoiding metaphysical questions: the less structure that is required to faithfully fix a term, the less metaphysical questions it involves. Terms such as *Red*, that need *more* structure than the standard model theory in order to be fixed faithfully, are less formal than those that can be fixed faithfully.

Thus, we obtain a measure of the formality of terms of some sort. The foregoing reasoning, however, does not give us a *criterion* for logical terms. The “most formal” terms would give us an extremely weak logic. Sher, McGee, Feferman and Bonnay limit generality in favor of other considerations, and draw the line before the logic is trivialized. I am not concerned here with where the line should be drawn, or whether indeed it should be drawn anywhere, but only at pointing to an underlying scale.

Our measure of formality connects the deep and the shallow senses of logical terms. The formality of terms, or, their logicity in the deep sense, is measured by their capability of being fixed in different structures, that is of what it takes for them to be logical terms

---

of operations  $K$ , let  $Sim(K)$  be the biggest class of functions between models leaving all operations in  $K$  invariant. Treated as a relation, let  $M(Sim(K)M')$  iff there is a function from  $M$  to  $M'$  in  $Sim(K)$ . Then  $Inv(\mathcal{F}) = Inv(Sim(Inv(\mathcal{F})))$ , and  $Sim(Inv(\mathcal{F}))$  is an equivalence relation (cf. (Bonnay, 2008, p. 40)).

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

in the shallow sense. And instead of a criterion for logical terms, we can propose the following principle: *the less structure a system needs for faithfully fixing its logical terms, the more metaphysics it avoids* (where “logical” is understood in the shallow sense). If reducing metaphysical assumptions is considered an virtue of a logical system, this principle recommends that we avoid fixing terms that require an elaborate structure.

Let us go back to some previous examples. We have seen that *Red* requires more structure than  $\forall$  if it is to be fixed faithfully. *Human*, on the other hand, does not require as much structure as *Red*: as a term denoting a necessary property (under our assumptions), it does not need a way to distinguish between worlds that have the same populations. However, it still requires more structure than  $\forall$ . *Human* will not be invariant under isomorphisms, and cannot be said to be determined by *card* in a thinned class of model, given some reasonable assumptions about the meaning of the term.

To end this chapter, I would like to bring the discussion back to the general framework of semantic constraints. In the two previous chapters, I argued that terms can be fixed in various ways and to different degrees, and that there need not be a strict dichotomy between the completely fixed and the non-fixed terms of a system. The discussion in this chapter centered on the terms of the system that are completely fixed. However, the same considerations can be utilized to deal with the more general question of which *semantic constraints* should be included in a logic. Invariance criteria gave us a way of evaluating the amount of structure needed to fix a term faithfully. Applying the lesson learned to semantic constraints, we shall speak of the amount of structure needed to *constrain* terms faithfully.

Fixing a term counts as a semantic constraint. We have seen that any constraint fixing *Red*:

$$I(\text{Red}) = \dots D \dots$$

in our original, accepted framework, will be unfaithful to its meaning. So such a constraint is ruled out, unless we are willing to add more structure to our model theory. But now take

## CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM

the constraint

$$I(Red) \cap I(Green) = \emptyset$$

that does not completely fix any particular term, although it does constrain the way *Red* and *Green* may be interpreted. There is no straightforward way we can apply invariance criteria on this semantic constraint. However, we may note that this clause constrains *Red* and *Green* *faithfully*. Taking on board that whatever is red all over is not green all over, this semantic constraint is in accord with the intended meaning of the terms, without imposing extra structure on the model theory. The situation is different with any semantic constraint of the form

$$I(Red) \cap I(Big) = \dots$$

where the only terms mentioned are *Red* and *Big*. It will either impose more structure than our standard model theory offers, or it will be unfaithful to the meanings of the terms involved, analogously to the case of just fixing *Red*. We can use invariance criteria as before to thin the model theory, and see which constraints can be made on terms that are faithful to their intended meanings.

We can now state a more general principle: *the less structure a system needs for faithfully constraining its terms, the more metaphysics it avoids*. Again, if reducing metaphysical assumptions is considered an virtue of a logical system, this principle recommends that we avoid constraining terms in a way that requires an elaborate structure.

### 8.4 Conclusion

The main criticism of invariance criteria for logical terms that I have presented in the previous chapter is that they assume the centrality of logical terms: that logical consequence is determined by the logical terms, and that there is a strict dichotomy between the logical and the nonlogical terms. However, invariance criteria are in accord with the proposed



## *CHAPTER 8. DETERMINACY, INVARIANCE AND FIXING A TERM*

view of formality as a means to avoid metaphysical questions. The invariance criteria that a term passes can be viewed as a measure for its logicity: as they tell how much structure is required to fix the term faithfully. The more structure is involved, the more metaphysical questions are expected to appear. But the idea behind this can be employed in the general framework of semantic constraints. Semantic constraints can be evaluated with respect to the amount of structure they require for faithfully constraining the terms they involve.

## Chapter 9

# Semantic Constraints: Modal Logic

The framework of semantic constraints can be naturally extended to incorporate modal logic. This chapter presents the basis for such an extension, mirroring two previous chapters relating to extensional logic. The first section consists of a modification of the framework to accommodate modal operators. In the second section I discuss what it is to fix a term in modal logic and give graded criteria for logical terms and for semantic constraints. Both parts of the chapter are aimed to show that the framework of semantic constraints is a good home for modal logic.

I will limit the discussion to propositional modal logic. The models of a language  $L$  will now be 4-tuples  $\langle W, w_0, R, v \rangle$  where  $W$  is a nonempty set of worlds,  $w_0 \in W$ ,  $R \subseteq W \times W$  an accessibility relation and  $v$  a valuation function for phrases in  $L$ .  $v$  is a binary function such that for each  $w \in W$  and phrase  $p$ ,  $v(w, p)$  is an element in the set-theoretic hierarchy with  $\{T, F\}$  as ur-elements. Given a set  $\Delta$  of semantic constraints for  $L$ , we say that a sentence  $\varphi$  in  $L$  is a logical consequence of a set of sentences  $\Gamma$  in  $L$  w.r.t.  $\Delta$  ( $\Gamma \models_{\Delta} \varphi$ ) if for any  $\Delta$ -model  $\langle W, w_0, R, v \rangle$ , if  $v(w_0, \psi) = T$  for every  $\psi \in \Gamma$ , then  $v(w_0, \varphi) = T$ .

Similar to what we had before, semantic constraints include universal quantification over models. We assume that the terms of  $L$  are:  $\{p_i\}_{i \in N} \cup \{\neg, \wedge, \vee, \rightarrow\}$  as well as some

modal operators. As before, each of the following constraints deals with a finite number of terms in the language (usually just one). A constraint that pertains to an arbitrary phrase of some sort (e.g. the first one below) should be regarded as a schema for infinitely many constraints, each referring to one phrase. And again, we assume the usual definition for *well-formed formula* (*wff*) for propositional logic.

## 9.1 Modal Systems

### 9.1.1 Standard modal systems

The set of constraints for standard modal logic includes the constraints for standard propositional logic, similar to 1-7 in Chapter 7. All the constraints we list have implicit universal quantification over  $w \in W$ .

1.  $v(w, p_i) = T$  or  $v(w, p_i) = F$  for all  $i \in \mathbb{N}$ .
2.  $v(w, \neg) = f_{\neg}$  where  $f_{\neg}$  is the negation function from truth-values to truth-values:  
 $f_{\neg} = \{(v_1, v_2) : v_1, v_2 \text{ are truth-values, } v_2 = T \text{ iff } v_1 = F\}$ .
3.  $v(w, \wedge) = f_{\wedge}$  where  $f_{\wedge}$  is the conjunction function from pairs of truth values to truth-values:  
 $f_{\wedge} = \{((v_1, v_2), v_3) : v_1, v_2, v_3 \text{ are truth-values, } v_3 = T \text{ iff } v_1 = v_2 = T\}$ .
4.  $v(w, \vee) = f_{\vee}$  where  $f_{\vee}$  is the disjunction function from pairs of truth values to truth-values:  
 $f_{\vee} = \{((v_1, v_2), v_3) : v_1, v_2, v_3 \text{ are truth-values, } v_3 = F \text{ iff } v_1 = v_2 = F\}$ .
5.  $v(w, \rightarrow) = f_{\rightarrow}$  where  $f_{\rightarrow}$  is the material implication function from pairs of sentences to truth-values:  
 $f_{\rightarrow} = \{((v_1, v_2), v_3) : v_1, v_2, v_3 \text{ are truth-values, } v_3 = F \text{ iff } v_1 = T \text{ and } v_2 = F\}$ .
6. For any wff  $s$ ,  $v(w, \neg s) = v(w, \neg)(v(w, s))$ .
7. For  $c = \wedge, \vee, \rightarrow$  and any two wffs  $s_1, s_2$ ,  $v(s_1 c s_2) = v(w, c)((v(w, s_1), v(w, s_2)))$ .

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

To the standard propositional portion of the logic we add constraints for the modal part, using the following definition:

**Definition 26** For each model  $M = \langle W, w_0, R, v \rangle$ ,

- (a) the function  $acc_M$  from worlds to sets of worlds is such that for each  $w \in W$   $acc_M(w) = \{w' : wRw'\}$ .
- (b) the function  $prop_M$  from wffs in  $L$  to sets of worlds in  $W$  is such that for each wff  $s$ ,  $prop_M(s) = \{w \in W : v(w, s) = T\}$ .

I omit the subscript  $M$  where context permits.

**Definition 27** (Modal operator) An  $n$ -ary *modal operator*  $O$  is a term such that for each model  $M = \langle W, w_0, R, v \rangle$  and  $w \in W$ ,  $v(w, O)$  is a function from an  $n$ -tuple of sets of worlds to truth values. (We will only deal with unary modal operators.)

- 8.  $v(w, \Box) = f_{\Box}^w$  where  $f_{\Box}^w$  is the box function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_{\Box}^w(X) = T$  iff  $acc(w) \subseteq X$ .
- 9.  $v(w, \Diamond) = f_{\Diamond}^w$  where  $f_{\Diamond}^w$  is the diamond function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_{\Diamond}^w(X) = T$  iff  $acc(w) \cap X \neq \emptyset$ .
- 10. For any modal operator  $O$  and wff  $s$ ,  $v(w, Os) = v(w, O)(prop(s))$ .

Note that truth-functional connectives can be construed as modal operators. For negation, the appropriate function would be  $\hat{f}_{\neg}^w$  defined by  $\hat{f}_{\neg}^w(X) = T$  iff  $w \notin X$ . Similarly, binary truth-functional connectives can be construed as binary modal operators.

If we would like to further constrain our modal logic to a specific modal system, we can add constraints on the accessibility relation  $R$ . These are semantic constraints in a broad sense: even though they do not explicitly constrain any term, they affect the way terms are interpreted. We can thus add any of the following constraints:

11.  $R$  is reflexive.
12.  $R$  is symmetric.
13.  $R$  is transitive.
14.  $R$  is euclidean.
15.  $R$  is universal.

So, for instance, adding constraints 11-13 or constraint 15 to 1-10 will give us  $S5$ . Constraints 11-15 do not mention any specific term, but they constrain the set  $\{\Box, \Diamond\}$  with respect to the constraints 1-10 (see Definition 5 in Chapter 7.3.1).

### 9.1.2 Other modal operators

Traditionally, modal logic deals with the terms  $\Box$  and  $\Diamond$  which in the most basic formulations are interpreted as operators of necessity and possibility. Playing with the accessibility relation  $R$  will vary the type of necessity or possibility dealt with. In some intensional logics, such as temporal logic, the intended readings of  $\Box$  and  $\Diamond$  are not “necessity” and “possibility”, but are still closely related, and the same constraints 8-10 can be used. The definition of a *modal operator* opens the door to operators that are different not only on account of  $R$ , but have a different kind of function associated with them. Analogous to generalized quantifiers, these are a result of a generalization of the standard modal operators.

We list some examples of such operators:

16.  $v(w, \Diamond^2) = f_{\Diamond^2}^w$  where  $f_{\Diamond^2}^w$  is the “double possibility” function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_{\Diamond^2}^w(X) = T$  iff  $|acc(w) \cap X| \geq 2$ .
17.  $v(w, \Box^-) = f_{\Box^-}^w$  where  $f_{\Box^-}^w$  is the “almost necessity” function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_{\Box^-}^w(X) = T$  iff  $|acc(w) \setminus X| \leq 1$ .

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

18.  $v(w, \Diamond^\infty) = f_{\Diamond^\infty}^w$  where  $f_{\Diamond^\infty}^w$  is the “infinite possibilities” function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_{\Diamond^\infty}^w(X) = T$  iff  $|acc(w) \cap X| \geq \aleph_0$ .
19.  $v(w, \Box^{-n}) = f_{\Box^{-n}}^w$  where  $f_{\Box^{-n}}^w$  is the “cofinite necessity” function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_{\Box^{-n}}^w(X) = T$  iff  $|acc(w) \setminus X| < \aleph_0$ .
20.  $v(w, \Box^{1/2}) = f_{\Box^{1/2}}^w$  where  $f_{\Box^{1/2}}^w$  is the “half necessary” function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_{\Box^{1/2}}^w(X) = T$  iff  $|acc(w) \cap X| = |acc(w) \setminus X|$ .
21.  $v(w, \Diamond^{-1}) = f_{\Diamond^{-1}}^w$  where  $f_{\Diamond^{-1}}^w$  is the “reverse diamond” function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_{\Diamond^{-1}}^w(X) = T$  iff there is a world  $w' \in X$  such that  $w \in acc(w')$ .
22.  $v(w, O^{Win}) = f_{O^{Win}}^w$  where  $f_{O^{Win}}^w$  is the “window” function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_{O^{Win}}^w(X) = T$  iff  $X \subseteq acc(w)$ . So  $O^{Win}s$  is true in  $w$  iff all worlds where  $s$  is true are accessible to  $w$  (see (Blackburn, de Rijke, & Venema, 2002, p. 426)).

The last two constraints, (21) and (22) are still in the spirit of standard modal logic, and the operators they define appear in textbooks (e.g. (Blackburn et al., 2002)). Constraints (16)-(20), on the other hand, make an unusual reference to the structure of possible worlds, and more specifically to the number of accessible worlds satisfying a sentence. Comparing the domains of models in extensional logic to sets of worlds in modal logic, the standard necessity and possibility operators are analogous to the standard quantifiers, and the operators in (16)-(20) are analogous to generalized quantifiers. It seems to me that such operators might have a place in a logical system, given some assumptions and a further understanding of the notion of a “number of possibilities”. For instance, modal logic is sometimes used to model the execution of a program. ‘ $\Diamond s$ ’ can be interpreted as: some terminating execution of the given program from the present state leads to a state bearing the information  $s$  (see (Blackburn et al., 2002, p. 12)). Generalized modal operators might have some significance

in that context, where the number of possible states in the execution of a program is taken into account.

The constraints I have given so far for modal operators fix them completely. We can also consider constraints for modal operators only fixing them to some extent, e.g.:

23.  $O$  is a modal operator.
24.  $v(w, O) \in \{f_{\Box}^w, f_{\Box-}^w, f_{\Box-n}^w\}$  (so given a set of worlds and an accessibility relation, there are three possible interpretations for  $O$ ).
25.  $O$  and  $O'$  are dual modal operators (see definition below).

I concentrate here on modal operators, but of course we can also fix to a certain extent propositional variables and sentential connectives using the apparatus of worlds and the accessibility relation.

### 9.1.3 Determinacy, dependence and categories

Some definitions of interesting properties and relations between terms in the framework of semantic constraints were presented in Chapter 7. Those can be naturally restated to fit the current modal framework.

**Definition 28** (Determinacy) Let  $a$  be a term and let  $B$  be a set of terms in  $L$ .  $a$  is *determined* by  $B$  w.r.t.  $\Delta$  if for any  $\Delta$ -models  $M = \langle W, w_0, R, v \rangle$  and  $M' = \langle W', w'_0, R', v' \rangle$  and every world  $w \in W \cap W'$ , if  $v(w, b) = v'(w, b)$  for all  $b \in B$ , then  $v(w, a) = v'(w, a)$ .

Analogously to Chapter 7, we add to any modal language the pseudo-terms  $\mathcal{W}$  and  $\mathcal{R}$ , and always include the constraints:

- $v(w, \mathcal{W}) = W$
- $v(w, \mathcal{R}) = R$

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

**Definition 29** (Logical term) A term  $a$  in  $L$  is a *logical term* (a *completely fixed term*) w.r.t.  $\Delta$  if it is determined by  $\{\mathcal{W}, \mathcal{R}\}$ .

Note that we chose the definition of *logical term* over other possibilities, e.g. having logical terms be determined by only  $\{\mathcal{W}\}$  or even  $\emptyset$ . This option will be discussed in the next subsection.

**Definition 30** (Dependency) Let  $A$  and  $B$  be sets of terms in  $L$ .  $A$  *depends* on  $B$  w.r.t.  $\Delta$  if there are  $\Delta$ -models  $M = \langle W, w_0, R, v \rangle$  and  $M' = \langle W, w_0, R, v' \rangle$  such that for any model  $M^* = \langle W, w_0, R, v^* \rangle$ , if for all  $a \in A$  and  $w \in W$   $v^*(w, a) = v(w, a)$  then for some  $b \in B$  and  $w \in W$   $v^*(w, b) \neq v'(w, b)$ .

The following definition is new, in the sense that it is not a modal reformulation of any previously given definition.

**Definition 31** (Dual operator) Let  $O$  be a unary modal operator.  $O$ 's dual,  $\tilde{O}$ , is the unary modal operator such that  $v(w, \tilde{O}) = f_O^w$  where  $f$  is a function from sets of worlds to truth values such that for  $X \subseteq W$ ,  $f_O^w(X) = T$  iff  $f_O^w(X^C) = F$ , where  $X^C = W \setminus X$ .

*Remark.* The dual relation is symmetric: if  $O$  is the dual of  $O'$  then  $O'$  is the dual of  $O$ .

**Examples** The following pairs of modal operators are duals:

- $\Box, \Diamond$
- $\Diamond^2, \Box^-$
- $\Diamond^\infty, \Box^{-n}$
- $\neg$  (construed as a modal operator) and the null operator  $+$  (with the accompanying function  $f_+^w$  such that  $f_+^w(X) = T$  iff  $w \in X$ )

**Claim.** Let  $O$  and  $O'$  be dual modal operators. Then:



1.  $O$  and  $O'$  are determined by one another.
2. If  $O$  and  $O'$  are not logical terms, then they depend on one another.

*Proof.* 1. Follows directly from the definitions of dual operators and determinacy.

2. By Proposition 3, Chapter 7, p. 138, adjusted to present terms.

□

The procedures of schematization and generalized schematization and the propositions regarding the validity of schemas is as in Chapter 7.3.4, and will not be reconstructed here.

## 9.2 Determinacy, Invariance and Fixing a Term: the Modal Logic Case

As I hope the previous section has shown, the framework of semantic constraints can be naturally extended to modal logic. The advantage of moving to the more general framework is three-fold. First, we have the general considerations presented in Chapter 6, urging for a view of formality that is not limited to a strict dichotomy between fixed and non-fixed terms, and allows fixing terms in different ways and to various extents. Accepting this view may lead to a host of new interesting and useful logical systems that are not limited to the conventions regarding logical terms. Secondly, the apparatus of the framework gives us useful tools to investigate the already existing standard as well as the generalized systems of modal logic. Thirdly, as I will argue in what follows, viewing a modal logic as a system of constraints imposed on a language, rather than viewing it through a strict division between logical and nonlogical terms, is more faithful to its nature.

The first two points were addressed in the previous section. I have only given a few examples of constraints that partly fix modal operators, but surely there are more examples

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

of constraints that might be of interest. Extending to first-order modal logic will allow constraints with richness close to that of Montagovian meaning postulates which are construed in his “intensional logic” (Montague, 1974c). Such extensions will not be presented in this work. Rather, I would like to give more substance the third point, and provide philosophical reasons for viewing standard modal logic from the perspective of semantic constraints.

This section extends ideas presented in Chapter 8. There I dealt with the question of what it is to fix a term in standard model-theoretic semantics. I also discussed the connection between invariance criteria for logical terms and what it takes to fix a term. Finally, I related the discussion to the perspective of semantic constraints. Here I shall confront special difficulties that arise when dealing with modal logic. I will conclude that even though these difficulties are not insurmountable, they do come at a cost which the framework of semantic constraints is discharged of paying.

The question of logicity in modal logic has various strands in the literature. Most of the work that has been done and the accounts that have been proposed are confined to extensional logic. But there are some exceptions. Peacocke, for example, uses the same criterion and considerations with the necessity operator as he does with extensional operators. His conclusion is that the necessity operator is ruled logical in the original formulation of his criterion, and nonlogical on a reasonable variation of it (Peacocke, 1976, p. 236f).

With respect to strictly extensional accounts, there are two ways to go when modal logic is considered: extending the account, or giving independent grounds for the exclusion of modal logic as acceptable logic. The latter approach is famously endorsed by Quine (e.g. (Quine, 1960, p. 195-200)). Representatives of the former include McCarthy, van Benthem, MacFarlane and Bonnay (McCarthy, 1981; van Benthem, 1989; MacFarlane, 2000; van Benthem & Bonnay, 2008). They all discuss ways of modifying invariance criteria to fit modal logic. The legitimacy of modal logic is an assumption I make here and do not attempt to justify. My main concern is how the discussions of logicity can be extended to

include modal logic.

### 9.2.1 Fixing terms and modal realities

As noted in Chapter 8, the question of which terms are logical (in the deep sense) goes hand in hand with the question of what it is to fix a term as logical (in the shallow sense) in a formal system. We shall see how these questions affect one another in the context of modal logic.

In previous chapters, I have distinguished between logical terms in the “shallow” sense—the terms that *are* fixed in a given system—and logical term in the “deep” sense—those that due to their distinctive properties *ought* to be fixed. The deep sense assumes the shallow sense: if some terms ought to be fixed, there needs to be some meaning to fixing a term. Transferring the question of logicality to modal logic, we must figure out what it is to fix a term in this setting before asking which terms should be fixed. In Section 9.1.3, Definition 29, I defined the logical terms in a system, i.e. the completely fixed terms of the system as those determined by  $W$  (signifying the set of worlds) and  $R$  (signifying the accessibility relation). However, I have given no grounds for that definition. Here I shall attempt to do that.

In standard systems, all modal operators are completely fixed in the sense of Definition 29. But in principle, any modal operator can be fixed or remained unfixed.<sup>1</sup> In standard logic, modal as well as non-modal, the grammatical category of a term is always fixed, that is, constant across all interpretations. The non-fixed terms are then those whose interpretation is allowed to vary completely freely within their grammatical category. In the framework I presented, even a term’s grammatical category need not be fixed. But in the usual sense of “non-fixed term”, a non-fixed modal operator  $O$  in the current framework is one which is constrained by items 23 and 10 (see above, and see also the definition of

---

<sup>1</sup>See (Kuhn, 1981) where the possibility of not fixing the modal operators is raised.

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

*modal operator*), and by no other constraint.

Recall the discussion in Chapter 8 on what it means to fix a term in the extensional case. Fixing a term didn't mean that its extension was to be identical in all models, but that its extension was to be a function of the domain. The thought was that the possible worlds represented by models vary, while the meaning of the fixed terms remains constant. The meaning is fixed *faithfully* when its interpretation is a function of the domains of models that is in accord with the term's intended meaning or intension: in each model, the term's extension represents the extensions it has in the possible worlds whose domains are represented by the model. Keeping the meaning constant and letting the world vary calls for a separation between the components of a model pertaining to the world and those pertaining to meaning. Something analogous is thus required for fixing terms in systems of modal logic.

Modal logic, however, requires us to reconsider the role of models and what they represent. Recall that the condition of *Necessity* was the reason we had models represent possible worlds in the extensional setting:

*Necessity*: For any valid argument  $\langle \Gamma, \varphi \rangle$ , necessarily, if all the sentences in  $\Gamma$  are true, then so is  $\varphi$ .

We used a possible worlds understanding of necessity, so that *Necessity* is understood as requiring, for any valid argument and any possible world, if the premises are true in the world, then so is the conclusion. *Necessity* is captured in the model-theoretic definition of logical consequence if every possible world is represented by some model. Specifically, we assumed, in the most part of this work, that domains of models represent domains (or *populations*) of possible worlds.

In modal logic, models seem to give us much more than a representation of some possible world: they give us an array of possible worlds, call it a (*modal*) “*reality*”.<sup>2</sup> Each

---

<sup>2</sup>The terminology here is confusing. The term *possible world* is used to denote both the metaphysical

model includes the members of a *frame*: a set of worlds and an accessibility relation. I attempt in this section, as in others in this work, to make as few assumptions as possible on possible worlds. I assume that necessity is truth in all possibilities, and use the term “possible worlds” to refer to those possibilities. I also assume that possible worlds can be represented mathematically by a set (representing their population) and some further semantic apparatus telling us which sentences are true at a given world.<sup>3</sup>

There are three salient possibilities for what frames might represent, that is, what the nature of “modal realities” might be. Some of them assume a different understanding of the condition of *Necessity* from the one we had in the extensional case.

1. *Frames represent higher order possibilities.* There are various ways things could have been. This is captured by a the “actual” frame. And then, there is the way *possibilities* could have been. Each possible range of possibilities is represented by a frame. We assume that each array of possibilities has a (“possible”) actual world. If this is the way we understand modal logic, then it calls for a strengthening of the condition of *Necessity*: an argument is valid only if in all possible arrays of possibilities where the premises are true, the conclusion is also true.<sup>4</sup> This approach is analogous to the approach in extensional logic that takes models to represent possible worlds, which includes both the metaphysical/representational approach taken from Etchemendy and to Shapiro’s blended approach (see Chapter 3). Of the three suggestions made here, this alternative is closest in spirit to the position I supported with respect to extensional logic in previous chapters (the blended approach), but it suffers from a contentious metaphysical assumption, or one might even say incoherence, referring to

---

object (or posit) represented by models, and to denote the items in  $W$ , which is part of the mathematically constructed model used in modal logic. These two uses are well-entrenched in the literature, so I will not provide alternative phrasing, and assume that each instance disambiguates by context.

<sup>3</sup>This way I am ruling out incomplete worlds, or situations (where not all sentences have truth values), as possibilities, mainly to keep things simple.

<sup>4</sup>A sentence is true in an array of possibilities if it is true in the actual world in that array.

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

higher-order possibilities.

2. *Frames represent the way possible worlds are.* That is, all frames represent the same “actual modal reality”. The variation between models, on this account, represents the variability in interpretations of the nonlogical vocabulary in modal languages. This option requires a revision of ordinary modal logic, so that there is only one frame, and only the valuation varies. This is analogous to the linguistic/interpretational approach (see Chapter 3). This approach can leave the condition of *Necessity* as it stands: when evaluating the validity of an argument we look at all possible worlds, but not at any higher order possibilities. But clearly, this approach cannot explain standard modal logic, where different frames are used in different models. There is no apparent reason for using different frames to represent the same array of possibilities.
3. *Frames represent higher-order epistemic possibilities.* We might think that there is only one “possible” array of possibilities, but we do not know how exactly it is structured. Or, we are not willing to commit to any one structure for fear of making contentious metaphysical assumptions. Thus, there may be various *epistemic* possibilities for the array of possibilities, and they are represented by frames. Here we can stay with the original condition of *Necessity*—it would still hold in this logic, as long as the “actual modal reality” is an epistemic possibility. The epistemic aspect and the avoidance of metaphysical assumptions are no longer an outcome of *Necessity*, and are more relevant to the condition of *Formality* as it was understood in Chapter 3: as a means to avoid unfounded metaphysical assumptions.

It seems to me that the third alternative, by which frames represent higher-order epistemic possibilities is most feasible. Now, our mission here is to fix the meaning of terms without fixing the “reality” (for lack of a better term) in which they are interpreted.

Note that *meanings* in this context can no longer be associated with intensions, the latter understood as functions from possible worlds to extensions. In general, the extension

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

of a modal operator is not determined by a possible world on its own, but by a whole array of possible worlds—usually, at the very least by those accessible to the world at which the operator is evaluated. Thus, meanings should be now associated with functions from “modal realities” to interpretations in those realities. Fixing a term, specifically a modal operator, in a *faithful* manner, amounts to giving it an interpretation that is in accord with its meaning.

The accessibility relation  $R$  poses a matter in question when we look into fixing terms.  $R$  was considered above to be part of “reality”. This means that it should not be held constant when a modal operator is fixed—since we are out to fix meaning, not reality. Indeed,  $R$  is a relation between worlds, and in that respect it is part of the reality of worlds. It is then used to determine the extension of modal operators—as other aspects of the given reality do. This is why in Definition 4 of logical terms in this chapter, logical terms (in the shallow sense) were defined as determined by the set including both  $\mathcal{W}$  and  $\mathcal{R}$ .

But then, on the other hand, there is reason to consider  $R$  as part of the meaning of modal operators, and thus fix it when modal operators are fixed. For instance, it might be part of the *meaning* of the  $S5$  modal operators that the accessibility relation involved has certain properties, such as being reflexive, symmetric etc. Indeed, in standard modal logic each system deals with one type of modality, and  $R$  is chosen to reflect a possible conception of that modality. In a generalized modal logic semantics, where multiple modal operators are allowed, a model will contain multiple accessibility relations. Thus,  $R$  is closely connected to the meaning of the modal operators: it is chosen so that the modal operators receive a correct interpretation.

Nonetheless, the last remark does not entail that  $R$  is part of the meaning of the modal operators, more than are other elements of reality that may affect their interpretation in a model. Even if modal operators with different meanings require different accessibility relations, those relations may still be regarded as part of reality: those are different aspects of modal reality that are relevant for different modal operators. So in this sense  $R$  should

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

not be fixed when we consider fixing a modal operator as a logical term.

In this perhaps artificial separation between meaning and reality, I thus take it that  $R$  is part of reality. More accurately, the various  $R$ s are part of reality. But the issue is not yet settled. We are aiming to fix only components of the model which pertain to meaning. But note that if we don't fix *any aspect* of  $R$ , we are barred from the wealth of modal systems that arises from various restrictions on  $R$ . In order to allow a variety of modal systems, we need a variety of accessibility relations, restricted in different ways. Perhaps the restriction of the accessibility relations relevant for the interpretation of a given modal operator is after all part of the operator's meaning? If so, specifying that accessibility relations relevant for  $S5$  operators are reflexive, symmetric and transitive, would be part of the meaning of the operators. We could then say that the ordinary  $S5$  semantics fixes the meaning of the  $S5$  modal operators without holding reality fixed.

Separating meaning and reality appears to be a hopeless task. Meaning comes into the picture when a connection is made between language and reality. I am unable to conclusively identify the point where it happens, and I doubt that it could be done in a non-artificial way. As will be argued towards at the end of the chapter, the system of semantic constraints is not committed to the separation between meaning and reality, and provides a platform for viewing the constraints on the interpretation of terms in a more holistic way concerning meaning and reality.

Say that we take all aspects of  $R$  to be part of reality, so it isn't affected by whether some term gets fixed. There is provisional way of maintaining the wealth of modal systems when considering the modal operators to be fixed.<sup>5</sup> We can look at the restrictions on  $R$  as minimal metaphysical assumptions needed for a logic of modality. And the more metaphysical assumptions we are willing to countenance, the more we can say about  $R$ . Even though we would like to avoid metaphysical questions to a large extent, we do not want

---

<sup>5</sup>The question of their logicity in the deep sense, and whether they *should* be fixed, will be dealt with shortly. What is on the table is whether they can, or in what sense they can, be fixed.



## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

to, nor can we, avoid all of them. We thus isolate a metaphysical assumption concerning  $R$  and take it on board, enriching our logic. Of course, we can relax or tighten these restrictions, where in one extreme  $R$  is any binary relation between worlds, and in the other  $R$  some predetermined relation between worlds—the same in every model (in such a case  $W$  would have to be fixed as well). In the former case, when we fix the necessity operator we obtain the system **K** of modal logic. In the latter case, where  $R$  is identical in all models we obtain either  $S5$  or a stronger system (depending on other features of the system).

To sum up the discussion so far, we have seen that fixing modal operators imposes a separation between the elements of a model pertaining to meaning and those pertaining to “reality”. The accessibility relation poses a particular challenge. Placing it on either side of the division between meaning and reality would not lead the standard systems of modal logic, where only some aspects of  $R$  are fixed. The restriction of  $R$  thus comes from considerations having to do with the metaphysical assumptions we are willing to countenance. As we shall see, the attitude towards  $R$  also influences which modal operators would come out logical in the deep sense by invariance criteria.

Before we move on, I would like to point out an advantage the framework of semantic constraints has in this discussion. To the extent that the idea of fixing a term in the system can be captured in the term-based framework, so it can be captured in the framework of semantic constraints. But the distinction between logical and nonlogical terms in the shallow sense, which is required by the term-based framework, is not required in that of semantic constraints. As we have seen, what it is to fix a term in the system is not a trivial question, even in the simpler case of extensional logic. We are required to separate the components of a model that represent meaning from those that represent reality. Recall that Shapiro, leaning on a Quinian rationale, circumvents the separation between meaning and reality by letting models (for extensional logic) represent possible worlds under interpretations of the nonlogical vocabulary, without distinguishing between two types of components of models (Chapter 3). But we learn that in order to speak about logical terms as those that are

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

or ought to be fixed, we are pressed to make this separation. Looking at things from the perspective of semantic constraints, we can fix certain aspects of the interpretation of a term without aiming to fix it completely. Moreover, semantic constraints can pertain to both meaning and reality without having to distinguish between the two.

The axiomatic historical origin of modal logic gives us a different perspective on modal operators. In an axiomatic system, the meaning of modal operators is conveyed by a set of axioms. However, it is imprecise to say that their meaning is completely fixed by the axioms. An axiomatic system perhaps captures the inferential behavior of a modal operator (or at least some of it), but it does not follow that the axioms give a full defined meaning. Each of the *S5* axioms gives us some information on the modal operators. In the semantic system we have constraints on *R* mirroring that information. There is thus no reason to think that the model-theoretic semantics for *S5* *completely* fixes the meaning of the modal operators.

Kuhn (1981) has suggested a semantical framework for modal logics, *operator logic*, where modal operators are purely schematic, that is, unfixed. Any restrictions on those operators comes in the form of *theories* (Kuhn, 1981, pp. 494f). This approach is in accord with the thought that modal logics do not give the full meaning of the modal operators, but present a thesis or a set of suppositions regarding them.

Kuhn's modal theories are comparable to meaning postulates, discussed in Chapter 6. The system, in its base, still has a distinction between logical and nonlogical terms, and some restrictions on the semantics are added onto the system. The framework of semantic constraints takes us a step further than Kuhn's approach. Kuhn's main motivation for constructing operator logic is to give a framework where modal operators can be treated as nonlogical terms, so that their logicity is no longer assumed by the formalism. The framework of semantic constraints, as well, includes the possibility of non-fixed modal operators. But moreover, by discarding the dichotomy between logical and nonlogical terms, it allows constraining them to various extents without fixing them completely. This means

that there need not be a defined boundary between meaning and reality. A given semantic constraint can pertain to both, without committing to a boundary.

### 9.2.2 Is *Box* a logical term?

Assume that we have settled on what fixing a term in a modal system amounts to. Let us now consider criteria for those terms that will eventually be fixed. Specifically, let us consider the possible extensions of invariance criteria to the modal logic setting. We can consult the literature on this issue: in modal logic, as in extensional logic, criteria for logical terms in the deep sense have gained some consideration, as opposed to the issue of logical terms in the shallow sense.

The idea is to translate invariance criteria to the modal setting. I shall consider three types of invariance criteria, employing permutations, isomorphisms or bisimulations. At first I shall refer mainly to permutation invariance, but most of the following remarks are relevant to isomorphism invariance as well. I will point out the differences later on.

In the extensional case, the permutations are on the domain of a model. The main question to be asked in the modal case is: What in the model is to be permuted (or, what structures do we compare to see if they are isomorphic)? In order to answer this, we need to answer another question: Which elements in a model can logical terms be sensitive to, and to which they should be indifferent? Proponents of invariance criteria refer to the generality and formality of logical terms (Sher, 1996; Bonnay, 2008), and characterize them as indifferent to how the world (or “reality”) is. This leads us back to the question of which parts of the model represent “reality”. Naturally, the set of worlds  $W$  should be part of what is permuted. The question confronting proponents of invariance criteria is whether  $R$  should be taken into account as well—namely, whether or not only permutations that respect  $R$  should be considered.

Now, obviously, logical terms, even if they are indifferent to what goes on in reality,

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

need not be indifferent to what pertains to their own meaning. If  $R$  is considered as a constituent in the meaning of the modal operators, rather than of reality, it is permitted to affect their interpretation. That is,  $R$  can justifiably pose a restriction on permutations on  $W$ . But it follows from our previous discussion, where the conclusion was that  $R$  is part of “reality”, that modal operators should not be sensitive to  $R$ .

Nevertheless, there might be after all aspects of reality which we would like to make assumptions about and allow to affect logical validity. In that case, logical terms would not be indifferent to *all* aspects of reality. So even if  $R$  is considered part of reality, there is the question of whether it is an aspect of reality we allow logical terms to be sensitive to or not. In other words, we return to the issue of what metaphysical assumptions we are willing to make with respect to  $R$ .<sup>6</sup>

Let us outline the results of the different possible approaches to  $R$ . We start with permutation-invariance. When permutations are at stake (as opposed to isomorphisms etc.) only one model is considered at a time. That is, an operator will be invariant under permutations of some (or all) models. First, consider all permutations on  $W$ , without fixing  $R$ .<sup>7</sup> As a result, the truth-functional connectives turn out to be logical. As for (other) modal operators: in the current setting, where the accessibility relation  $R$  is allowed to vary freely,

---

<sup>6</sup>MacFarlane refers to the features of models to which logical terms are allowed to be sensitive as *intrinsic structure*. See discussion in (MacFarlane, 2000, pp. 219-237). MacFarlane’s Principle of Intrinsic Structure (MacFarlane, 2000, p. 224) deems  $R$  not a part of intrinsic structure, thus it should not be a fixed part of models. The Principle says: “One should use a type with just as much intrinsic structure as is needed for the postsemantics, and no more.” (*ibid*). By *postsemantics* MacFarlane refers to the considerations involved in assertion and inference. Those, in turn invoke semantic notions, such as truth and falsity. In the case of modal logic, according to MacFarlane, the accessibility relation is not *required* for the postsemantic level, and therefore should not be invoked. This approach is absolutist regarding logicity (whereas mine is *gradualistic*) and language dependent: for a given language there is a strict boundary between logical and nonlogical terms.

<sup>7</sup>Since we are mainly interested here in modal operators, we can stay with the propositional part of modal logic when testing the different criteria.

only those operators that are indifferent to  $R$  will be invariant. Recall that modal operators were defined by functions from sets of worlds to truth values. A permutation-invariant operator  $O$  defined by the function  $f_O^w$  (where  $w$  is the world of evaluation) is such that for every  $w, w' \in W$  and every  $X \subseteq W$ ,  $f_O^w(X) = f_O^{w'}(X)$ . So, for instance, the necessity operator  $\Box S5$ , as defined by  $f_{\Box S5}^w(X) = T$  iff  $X = W$ , is permutation invariant under our assumptions. Note that this is a way of defining the  $S5$  box without reference to  $R$ . The permutation invariance of the necessity operator defined with reference to  $R$  as in constraint (8) will depend on  $R$ .

Note, again, that we always permute a set of worlds of some specific model with a given accessibility relation. The given model is a parameter in the condition of permutation invariance: the results are relative to the model chosen (just as Tarski's permutation invariance condition in extensional logic is domain relative). So the results of the permutation invariance criterion will depend on what  $R$  is in the given model. The  $S5$  necessity operator can be defined by constraints (8) and (15): the latter restricting  $R$  to be universal. Defined this way, the  $S5$  box will still be permutation invariant. Indeed, if the accessibility relation in this model is universal or empty, the necessity operator defined by (8) is permutation invariant. Otherwise, apart from some trivial cases of  $W$ , it isn't. Note that including constraints (11)-(13) (that constrain  $R$  to be reflexive, symmetric and transitive) is yet another way to obtain the  $S5$  necessity operator, but defined thus it is not permutation invariant. See (MacFarlane, 2000, p. 217).

Now there is a difference between permutation of just one model and isomorphism between different models. The isomorphism invariance condition is not limited to dealing with one model at a time. Instead, we consider the class of all models and compare the extensions of terms in those that are isomorphic. Since we are not having  $R$  fixed at the moment, the relevant isomorphisms to look at are between sets of worlds (i.e. bijections) not necessarily respecting  $R$ . Two models will be considered isomorphic if there is a bijection between their sets of worlds. According to the invariance criterion received,  $\Box S5$  as defined

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

on the previous page will be logical, but there will be no case for the box operator as defined in (8) to be logical.

To conclude this case where  $R$  is not constrained, only modal operators that are independent of  $R$  (are not defined with respect to  $R$ ) will in general pass the invariance criterion. In the specific case of permutation invariance relative to models where  $R$  is universal or empty, or  $W$  is trivial, other modal operators will also pass the test. The  $S5$  modal system turns out to have a special status. They can be defined without reference to  $R$ , or with reference to a universal  $R$ , and thus defined they are permutation invariant. This result has led some authors to consider  $S5$  operator as the only logical modal operators (McCarthy, 1981; MacFarlane, 2000).

Now assume that we are taking upon ourselves the metaphysical commitments bound up with constraining  $R$ . That is, we are limiting in advance the class of admissible models to those where constraints such as (11)-(15) are satisfied. We also consider only permutations that respect  $R$ . In that case we allow only permutations  $\pi$  on  $W$  such that for each  $w, w' \in W$   $wRw'$  iff  $\pi(w)R\pi(w')$ . This will make  $\Box$  as defined in (8) a logical term, regardless of restrictions on  $R$ . It would also rule the non-standard operators listed in (16)-(22) as logical. The results are the same when looking at isomorphisms. As in the case of generalized quantifiers, we can strengthen the invariance criterion to rule out the non-standard modal operators. This can be done by considering, instead of permutations

or isomorphisms, bisimulations between models.<sup>8</sup> Bisimulations match more closely the standard systems of modal logic. Further, van Benthem and Bonnay (2008) argue that bisimulations are the modal logic correlates of potential isomorphisms, and using them in defining logicality for their respective systems is based on similar considerations.<sup>9</sup>

### 9.2.3 Modal logic and degrees of logicality

In the previous discussion we distinguished between the components of a model to which a term is sensitive to, and those to which it is indifferent. Invariance criteria tell us which components should be on each side of the line when it come to logical terms. Now, the set of elements in the model to which a term is sensitive *determine* it, in the sense defined and discussed in Chapter 8.

As we have seen in the previous chapter, what determines a term is the amount of structure needed for fixing it in a manner faithful to its meaning. When discussing extensional logic, I proposed the following principle, applicable also here:

The less structure a system needs for faithfully fixing its logical terms, the more metaphysics it avoids (where “logical” is understood in the shallow sense).

---

<sup>8</sup>A bisimulation  $BiS$  between two models  $M$  and  $M'$  for propositional modal logic satisfies the following conditions:

1.  $BiS$  is a relation between  $W$  and  $W'$  such that if  $w \in W$  and  $w' \in W'$  and  $wBiSw'$ , then for any propositional term  $p$ ,  $v(w, p) = v'(w', p)$ .
2. For each  $w \in W$  and  $w' \in W'$  such that  $BiS(w) = w'$ , for each  $w_* \in W$  such that  $wRw_*$  there is a  $w'_* \in W'$  such that  $w'Rw'_*$  and  $BiS(w_*) = w'_*$ .
3. For each  $w \in W$  and  $w' \in W'$  such that  $BiS(w) = w'$ , for each  $w_* \in W$  such that  $w_*Rw$  there is a  $w'_* \in W'$  such that  $w'_*Rw'$  and  $BiS(w_*) = w'_*$ .

See (Blackburn et al., 2002, p. 64f).

<sup>9</sup>van Benthem and Bonnay show the connection between bisimulations and potential isomorphisms by proving that potential isomorphism is the coarsest invariance relation preserving  $\exists$ , and bisimulation is the coarsest one preserving  $\Diamond$  (van Benthem & Bonnay, 2008).

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

I have suggested that the condition of *Formality* is motivated by the aspiration to avoid as much metaphysics as possible. *Formality* thus discourages us from fixing terms whose meanings demand an intricate structure when modelled. Now modal operators demand a much more involved structure than there is in extensional logic. This might deter some, but in the end it is a question of balance between expressive and argumentative power on the one hand and metaphysical safety on the other.

I have not cast doubt on the minimum added structure imposed by including the category of modal operators, namely the set  $W$  of possible worlds. I have rather concentrated on the accessibility relation between worlds. The non-*S5* operators need this extra structure, and therefore commit us to more metaphysics than an operator indifferent to  $R$ . And once accepting sensitivity to  $R$ , any restriction on it leads us to some further assumptions.<sup>10</sup>

Invariance criteria for logical terms, extended to the modal case, give us a way of “measuring” the amount of structure needed by some terms. Take the criterion of isomorphism invariance, which can be construed as sensitive or as insensitive to  $R$ . In either version, it rules out modal operators that are sensitive to the identity of particular worlds. The criterion of bisimulation invariance is sensitive to  $R$ , and is more restrictive than the sensitive-to- $R$  version of the isomorphism invariance criterion. It rules out operators that are sensitive to features of  $W$  and  $R$  such as the number of worlds accessible from some world. In this sense we can say that the operators defined by constraints (8)-(9) are more formal than those in (16)-(20), which require a more elaborate understanding of necessity, possibility, and accessibility between worlds. The *insensitive-to- $R$*  version of the isomorphism invariance criterion is stricter than the sensitive-to- $R$  version, but is neither stricter

---

<sup>10</sup>One might be concerned that by talking about *metaphysical* questions I am already assuming a boundary between meaning and reality. I should stress how freely I use the term *metaphysics*, possibly reflecting my ignorance in that subject. By using this term I refer to an extensive and intricate body of knowledge that includes matters of both reality and meaning. I am in no way assuming a boundary between reality and meaning, at the risk of being guilty of abuse of the term *metaphysics*.



nor less lenient than the bisimulation invariance criterion. We should remember that the talk of “degrees of structure” is loose, assuming no linear order. This also means that invariance criteria giving us a “scale” for the logicity of terms should be understood accordingly. In the end, what matters is isolating the metaphysical assumptions that we are willing to make, and by that deciding on the acceptable structures in our model theory.

Again, we can extend this idea beyond the term-based approach, and employ it in the context of semantic constraints. We learn from invariance criteria which structures are needed for fixing various terms, and we can use a similar line of reasoning for semantic constraints in general. We move to the more general principle presented in the previous chapter:

The less structure a system needs for faithfully constraining its terms, the more metaphysics it avoids.

The “degrees” of structure are set by the similarity relations used by invariance criteria. This idea was spelled out in the previous chapter, and I shall not repeat the details. Suffice it to say that just as fixing a term faithfully requires some amount of structure, so does a semantic constraint which does not completely fix a term, but is still faithful to the meaning of the terms it contains.

### 9.3 Conclusion

Modal logic poses specific challenges when posed with the issues of what it is to fix a term, and which terms ought to be fixed. The accessibility relation presents difficulties in the consideration of both issues. Given that fixing a term requires a separation between the components of models representing reality and those representing meaning, we have seen after some consideration that if such a separation is carried out,  $R$  should belong to the former. Thus, fixing modal operators in the logic does not entail that  $R$  should be fixed.

## CHAPTER 9. SEMANTIC CONSTRAINTS: MODAL LOGIC

Yet, there are other considerations that would have us fix some aspects of  $R$ , depending on the metaphysical assumptions we are willing to make when constructing a modal system.

We are confronted again with the dilemma of  $R$  when looking at invariance criteria for logical terms. We need to decide whether logical terms should be sensitive to the structure imposed by  $R$  on  $W$ , and this again comes down to what metaphysical assumptions we are willing to make with respect to  $R$ . Added metaphysical assumptions are marked by added structure: terms that need the added structure to be faithfully fixed are more metaphysically loaded. The invariance criteria that incorporate a given metaphysical assumption stand in a lower degree of logicity, *mutatis mutandis*, than those that don't (and not assuming there is a linear order of degrees of logicity). The same line of thinking applies also to the general framework of semantic constraints.

Finally, we can appreciate the added value of the perspective provided by the framework of semantic constraints. Section 9.1 showed that the accommodation of modal logic in the framework is technically feasible and promises fruitful extensions. In addition, some of the difficulties encountered in Section 9.2 can be avoided in a semantic constraints framework. This framework does not depend on a dichotomy between logical and nonlogical terms, neither in the shallow nor in the deep sense. It is therefore not committed to a separation between reality and meaning. We can make various restrictions on models and impose constraints *vis-à-vis* some terms without claiming to have fixed their meaning with respect to a varying reality. The restrictions imposed through semantic constraints need not be categorized as either pertaining to meaning or pertaining to reality, and thus a line need not be drawn dividing those two realms. Instead, we can take a more holistic stance. We can focus on the nature and extent of metaphysical assumptions we are willing to accept in our logic—"metaphysical" pertaining to both reality and meaning—and from there construct models through the appropriate semantic constraints.



## Part III

# Logic and Language



## Chapter 10

# Logical Consequence and Natural Language

The relation between logic and natural language is one of the most intriguing issues in philosophy. It is indeed the driving motivation of this work. Nonetheless, the most part of the dissertation deals with formal languages. In the nature of things, research of what are supposed to be just preliminaries, takes on a life of its own, and ends up consisting the major part of the work. So the subject of logic and natural language, with all its complications and intricacies, cannot be given comprehensive treatment in this work. This chapter, therefore, is more of an initial effort in an extensive project.

The following chapter consists of three sections. The first section is the main one, in which I ask whether there is a logical consequence relation in natural language. I spell out what I mean by logical consequence and its being (or not being) in natural language, and, ultimately, give a negative answer to the question. I first diagnose two projects prevailing in the literature that have to do with the question. One is the traditional project whose aim is to develop rigorous methods for scientific reasoning, and whose most notable representative in modern times is Frege. In addition, there is the more recently developed linguistic project,

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

which uses the tools of logic (formal systems etc.) for the mostly empirical investigation of natural language. My own orientation throughout this work has been traditional: I was not concerned with characterizing phenomena in natural language, but with analyzing philosophical concepts having to do with logical consequence.

Natural language, I assume in the discussion, is part of the natural world, and is to be studied through empirical science, as is done in the linguistic project to various extents. Thus, the question of logical consequence in natural language becomes: Can the empirical linguistic project provide the traditional project with a satisfactory relation of logical consequence? My answer is no. The argument rests on assumptions I make regarding what the traditional project needs, and regarding what the empirical project can provide.

In the second and third sections I discuss some of the consequences of the view, and how it connects with positions taken in the rest of the work. In the second section I provide a positive stance towards the connection between logic and natural language. Even though logic is not to be found in natural language as it is, it can be imposed on language through the process of formalization. I propose to look at logic as a system of commitments made with respect to language, a view that is consonant with the framework of semantic constraints developed in Part II.

In the third section I revisit the logic-as-model approach that was employed throughout this work. According to this approach, formal systems model logical consequence in natural language. But if there is no logical consequence in natural language, what is there to model? I claim that still various parts of formal systems can be seen to represent various elements in the world and in natural language (e.g., models represent possible worlds), so that some of the virtues of the logic-as-model approach can be retained.

## 10.1 Is there logical consequence in natural language?

The question in the title is difficult to answer not just because of the intricate issues involved, but also simply because it is not exactly clear what it means. I will try to give some substance to the question by explaining the concepts involved. That is, I will have to explain what I mean by “logical consequence” (this will rely on previous chapters), by “natural language” and finally by “natural language having logical consequence”. I will also discuss the notion of *formality*, the feature of logical consequence I will mainly focus on (Section 10.1.2).

The question of logic and natural language is deeply rooted in the analytic tradition. It would be instructive to look into the attitudes expressed in the literature. In some cases, difference in attitude is an outcome of a different use of concepts and of what it means for natural language to have a logic. In the first section (10.1.1) I survey some of the main views expressed on the subject. I will be concerned with modern conceptions of logic, where formal languages play a prominent role.

I will devote much attention to how authors view the relation between natural and formal language. It is often assumed that formal languages are correct vehicles for logical consequence, and thus their similarities and contrasts with natural language can shed light on the question of logical consequence in natural language.

In the historical survey, I will distinguish between two relevant projects and their different aims: the traditional project of logic and the linguistic project. The former is the age-old logical-philosophical enterprise of defining validity for the sake of reasoning in science. The latter project is that of linguistic study which uses formal tools of logic in the investigation of natural language. As might be expected, these two projects provide different perspectives on the relation between logic and natural language.

Finally, I will be able to answer the question at hand. I will adopt the perspective of the traditional project, as I have done implicitly in previous chapters. I will show that the con-



## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

ceptual assumptions made in Section 10.1.2 together with minimal principles extracted from the traditional project lead to the conclusion that there is no relation of logical consequence in natural language.

### 10.1.1 Making sense of the question: historically

In modern philosophy of logic, the attitude towards natural (colloquial, ordinary) language in many cases stems from the contrast with formal or formalized languages. Frege famously compared his ideography (*Begriffsschrift*) to a microscope, and ordinary language to the naked eye. Each has its own advantages. Frege's ideography, as the microscope, is better suited for scientific purposes. The microscope provides sharpness of resolution, and likewise, Frege's ideography is more accurate than ordinary language inasmuch as it is free from the misconceptions that arise through the use of ordinary language (Frege, 1967, pp. 6f). Frege thus views his ideography as a contribution to the methodology of science, a point to which I will return later on. For Frege, logical relations and truth ultimately concern thought (or conceptual content), which is non-linguistic (*ibid*, see also (Frege, 1984)). Ordinary language and his ideography are both concrete means of expression—both are used to express thoughts—and the latter makes a better tool for the investigation of ideas. The main difference is that the ideography presents the logical structure of thought in a perspicuous manner, so that inferential relationships become transparent (*cf.* (Frege, 1979, pp. 12f), (Frege, 1979, p. 252), see also (Bar-Elli, 2004, p. xxxix)). In Frege's case the issue is not whether natural (ordinary) language has logical consequence, since logic primarily belongs to the realm of thought. The question is rather whether natural language can be used as a secure means of expression of inferences. Frege's answer to the question thus reformulated is negative.

In the same vein, Russell has acknowledged the vagueness and ambiguity in natural language. However, as was exemplified in his theory of definite descriptions (Russell, 1905),

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

he contended that sentences of natural language have a logical form that can be revealed by formal means.

The so-called imprecisions of natural language have been a reason for logicians after Frege and Russell to abandon natural language in logical investigations. Frege's ideography, transformed into modern-day first or second order predicate logic, provides a most effective alternative. Tarski, for instance, stresses that a structural characterization of language is essential for a structural definition of truth, and thus also of logical consequence, as it leans on the definition of truth. But this is hopeless when natural language is concerned:

For this language is not something finished, closed, or bounded by clear limits. It is not laid down what words can be added to this language and thus in a certain sense already belong to it potentially. We are not able to specify structurally those expressions of the language which we call sentences... *The attempt to set up a definition of the term 'true sentence'—applicable to colloquial language is confronted with insuperable difficulties.* (Tarski, 1933, p. 164)

In his discussion on the concept of logical consequence, Tarski remarks on the features of everyday language, and explains why the concept cannot be defined in complete adherence to common usage:

With respect to the clarity of its content the common concept of consequence is in no way superior to other concepts of everyday language. Its extension is not sharply bounded and its usage fluctuates. Any attempt to bring into harmony all possible vague, sometimes contradictory, tendencies which are connected with the use of this concept, is certainly doomed to failure. (Tarski, 1936, p. 409)

Going back to his discussion of truth, Tarski contrasts between colloquial and formalized languages:

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

Whoever wishes, in spite of all difficulties, to pursue the semantics of colloquial language with the help of exact methods will be driven first to undertake the thankless task of a reform of this language. He will find it necessary to define its structure, to overcome the ambiguity of the terms which occur in it, and finally to split the language into a series of languages of greater and greater extent, each of which stands in the same relation to the next in which a formalized language stands to its metalanguage. It may, however, be doubted whether the language of everyday life, after being ‘rationalized’ in this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalized languages. (Tarski, 1933, p. 267)

The critical obstacle for a logical treatment of natural language, according to Tarski, is its being semantically closed. Since natural language contains its own truth-predicate, a liar sentence can be constructed, and it is faced with paradox (Tarski, 1933, pp. 154-165). It is not clear what would be Tarski’s position towards treating logically a non-semantically closed fragment of natural language. The above quotations imply that ambiguities and the overall unstructuredness of natural language are serious impediments to such a treatment.

Carnap has also noted the “unsystematic and logically imperfect structure of the natural word-languages (such as German or Latin), the statement of their formal rules of formation and transformation would be so complicated that it would be hardly feasible in practice” (Carnap, 1937, p. 2). Carnap, as Tarski, has generally restricted his theories to formalized languages ((Carnap, 1963b, p. 918), see also (Carnap, 1963a, p. 931)). Nonetheless, in his reply to Beth in the Schilpp volume, Carnap agrees that logical analysis of natural languages is possible in the following senses: “(1) an empirical description of the most important and most frequently used syntactical forms occurring in natural language, with indication of their frequencies, but without any claim of completeness; or (2) the complete representation of the syntactical structure of a constructed language which is to some extent similar to the

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

syntactical structure (e.g., order of words) of a part of a certain natural language.” (Carnap, 1963a, p. 931). Thus, despite Carnap’s scepticism regarding the possibility of a complete logical description of natural language, a partial one, describing predominant phenomena in natural language, is feasible—as in his (1). Another way to go about the issue of logic in natural language is to use an artificial language which can be described in a complete manner, which can be said to resemble a fragment of natural language (2). Carnap notes that the investigations of the kind of (1) are different from those of constructed languages, as they are empirical and less exact. Investigations of kind (2) are usually carried out by logicians rather than linguists.

One can conceive of a formal language as a regimentation of ordinary language. Quine suggests that regimentation is a means to resolve ambiguities and simplify logical theory (Quine, 1960, §33). In the process of regimentation, sentences of ordinary language are paraphrased into logical notation. The formal sentence which is the outcome should not be required to be synonymous with its origins, since it is non-ambiguous while the origin might be ambiguous. The formal sentence is rather a simplified replacement for the original one. For Quine, it seems, the difference between formal and natural language is a matter of degree of simplicity and perspicuity. According to Quine, formal language is preferable for the construction of logical or scientific theory; in this point Quine is in agreement with Frege, Tarski and Carnap. What distinguishes Quine, especially from Frege, is his empiricist outlook and his rejection of the notions of analyticity and the *a priori*. The choice of a logical system, however, is not made on a solely empirical basis by Quine, but rather he settles on standard first-order logic through theoretical and arguably pragmatic considerations (Quine, 1986, p. 79). Thus, as will be explained later, logic for Quine is not a phenomenon in natural language, inasmuch as systematic virtues, over and above mere empirical data, have a decisive role in settling the right logic.

More recently, Azzouni claimed that natural language, and specifically English, is inconsistent. He bases this on inferences where, he claims, each step would be intuitively

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

acceptable to any speaker of English. The inferences are essentially in the form of Curry's paradox (Curry, 1942). For any sentence whatsoever there is such an inference with the sentence as its conclusion. So in ordinary (unregimented) English, any sentence can be inferred. From this Azzouni draws the conclusion that the logic of (unregimented) natural language is trivial: every sentence is both true and false (every sentence and its negation are derivable from acceptable assumptions, thus both are true, thus non-negated sentence is both true and false). Azzouni then claims that inconsistencies can be and indeed are avoided in *regimented* English (Azzouni, 2003, 2006, 2012). Henceforward I will leave paradoxes aside, and will not consider them in later discussion.

By contrast to the overall negative approaches to the question of logic and natural language, we can find a positive attitude in approaches to logic with a linguistic orientation, which place natural language as the primary subject of inquiry. The working hypothesis in such studies is that natural language has its own logic, and there is a task of finding it out. Hence, one can speak of the logic of certain locutions in English such as quantifiers and conditionals through a formal description of their function in truth conditions or inference or entailment relations. Thus, thinkers in the semantic tradition of the mid 20th century onwards speak of the applicability of formal language as a tool in the study of natural language (*cf.* (Davidson, 1967, p. 313)). The change from a negative to a positive perspective on the question of logic and natural language will play an important role in the discussion later on in this chapter.

Davidson is a central representative of the positive approach. His aim was to construct a theory of meaning to natural language, which involves an empirical study of language and the use of formal methods taken from Tarski's theory of truth (Davidson, 1967, 2008). In the positive semantic approach, usually attributed to Davidson and Harman, formal languages are seen as representing underlying structures in natural language. More specifically, the logical form of a sentence – the feature of the sentence that determines its truth-conditions or inferential character – is exhibited by a formula in a formal language conceived of as

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

a *canonical idiom*. These formulas are sometimes termed *semantic representations* of the associated natural language sentences (see (Lycan, 1984, p. 13)). Furthermore, linguistic studies in the 1960's made increasing use of formal logic, which led to the identification of the *deep structure* of a sentence with its logical form (Lycan, 1984, pp. 5f), (Harman, 1972a; Davidson, 2008).

The development of the notion of transformational grammar reinforced the positive attitude towards logic in natural language. However, Chomsky himself, who advanced and promoted the usage of formal tools in linguistic theory (Chomsky, 1957), had reservations about relying on such methods in the linguistic study of synonymy or consequence relations (Chomsky, 1955): he claimed that these notions should be studied empirically, and the success of such research would be unaffected by the employment of logical tools (*ibid*, p. 45).

Montague, in a different branch of the study of semantics, is a notable representative of the positive approach (Montague, 1974a, 1974c, 1974b). Montague has claimed that there is no important theoretical difference between formal and natural languages ((Montague, 1974a, p. 188), (Montague, 1974c, p. 222)). The syntax and semantics of both kinds of languages, Montague's work aims to show, can be comprehended within a single natural and mathematically precise theory (*ibid*). Montague presents a fragment of English through a division of basic expressions into semantic and syntactic categories, and a treatment of complex expressions through recursive semantic and syntactic rules, in the framework of intensional logic. Logical truth and logical implication for English are defined in the Tarskian model theoretic way. The complications previous authors have alluded to: ambiguities and intricate grammatical phenomena in natural language are handled by Montague with formal means.

Montague's claim regarding the relation of formal and natural language could seem too strong, especially if artificiality is considered an essential feature of formal languages. Natural language, viewed as a natural phenomenon, thus belongs to a different category from

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

constructed formal languages. Taking this distinction into account, Montague's contention could be modified to be that there is a mechanical translation from natural to formal language.

Another statement of the relationship between formal and natural language is taken from the natural sciences, where *models* are used to study natural phenomena. Formal languages may be conceived as models for various phenomena in natural language. Such an approach is intimated in (Chomsky, 1957, p. 5), (Chomsky, 1980) and to some extent in (Davidson, 2008). Specifically, as Shapiro suggests, formal languages (or more accurately, formal *systems*) may be seen as models for logical consequence in natural language (Shapiro, 1998). This view of *logic-as-model* was presented in Chapter 3, and will be revisited in Section 10.3.

As we can see so far from this short survey, authors fall roughly into two very different projects dealing with the question of logic and natural language. In the *traditional project*, logic assumes its traditional role as a methodology for the sciences. Frege, Tarski and Carnap, each in his own way, are modern representatives of this approach. By contrast, the *linguistic project* is concerned with the investigation of natural language with the use of formal means. Natural language is assumed to have a logical consequence, and the task of the linguistic project is to characterize it. It is an empirical project, insofar as language is considered as a natural phenomenon. The formal languages devised in the traditional project have thus found their way to the linguistic project, but have changed considerably in their new position. The simple, yet solid, first or second-order languages employed by Frege and his predecessors proved insufficient for the study of natural language with the intricate structures it exhibits. Yet simplicity and moreover restrictiveness are of crucial importance for the epistemological underpinnings of the traditional project.<sup>1</sup>

We should not see the linguistic project as superseding an old tradition. The traditional

---

<sup>1</sup>The logicist program is a manifestation of the approach to logic as a tightly bounded discipline, for if it were not such a discipline there would not be much interest in reducing mathematics to it.

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

project, that seeks a method for correct reasoning is still prominent in philosophy (at least the philosophy of logic and mathematics), as can be seen in textbooks (Copi, 1961; Sainsbury, 1993) and the research projects discussed in Part I of this work.

The traditional project is strongly connected to epistemological motivations. For Frege, for instance, logic is part of a theory of ideal justification (Frege, 1964). Wagner (1987), considering Frege specifically, describes the *rationalist conception of logic*, which aims at perfect rigor and ideal systematization of knowledge. The epistemological aspects of Frege's conception are at odds with the empiricist members of the traditional project, most notably with Quine. The main divide concerns the notion of the analytic, basic in Frege's conception and unacceptable for Quine. However, both Frege and Quine view logic as an instrument for the systematization of science, and this is the aspect of the traditional project that I will focus on.

The linguistic project and the traditional project can be viewed as complementing each other, a combination that would be especially appealing from an empiricist perspective. The latter is normative, inasmuch as it is concerned with method. The former is a descriptive project, that can provide the basis for the construction of the method for the latter. The advantage from the empiricist viewpoint is that rationalist methods characteristic of the traditional project, such as relying on intuitions of the *a priori*, are replaced with an empirical study of language. This can work only if study of natural language can in fact provide us with the right tools required for the traditional project.

In fact, the two perspectives to logic meet in recent literature, though phrased in different terms from those I have been using. The question bringing them together is: Does natural language present us with a logical consequence relation of the kind prevailing in the traditional project? As we have seen in Part I, the distinction between logical and nonlogical terms is a central issue in the traditional project. Thus, one can ask more specifically: Does natural language present a distinction between logical and nonlogical terms? That is, is there a category of logical terms that arises from the investigation of natural language?



## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

This question is asked and answered negatively by philosophers of language and logic such as Harman, Lycan and Glanzberg. Let us consider their positions in turn.

Harman (1984) looks for a logical consequence relation determined by grammar. He is interested in ordinary reasoning, as conducted in natural language. Harman concludes that grammar does not provide a relation of logical consequence. Harman assumes that “Logical implications hold by virtue of logical form, where this is determined by the grammatical structure and logical constants involved.” (p. 120). To this he adds the following two assumptions regarding logical constants: “The logical constants are grammatically distinctive in a way that has something to do with their special role in reasoning.” (p. 121) and “The logical constants are those words which belong to small closed classes of atomic members of logical categories.” (p. 122). Another assumption he adds is: “All immediately intelligible implications expressed in language are logical implications” (p. 124). However, according to Harman these assumptions collapse when applied to natural language. In natural language we find a large category of immediate implications, such as “*S* knows that *P*, so *P*” and “It is true that *P*, so *P*” (p. 124). These implications depend on the words involved, and would not hold were “knows” or “true” replaced with other words from the same grammatical category. But, according to Harman, “knows” and “true” are not logical constants, as they are best treated as members of large open classes of atomic members of some logical category. “Know” should be classed with “believe”, “hope”, “fear”, “expect”, “regret”, and so on, which are not logical constants. And “true” should be classed with “unlikely”, “probable”, “possible”, “delightful”, “surprising”, and so on’ (*ibid*). Harman considers the possibility that the examples of immediate implications given have hidden assumptions, but gives further argumentation against this option (p. 125). For my purposes it suffices to note that Harman reaches his negative conclusion regarding logic and natural language through assumptions regarding the role of logical terms in logic. I will consider this mode of argumentation in the next section, when giving my own answer to the the question of

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

logic and natural language.<sup>2</sup>

Lycan (1989) claims that natural language does not display a strict dichotomy between logical and lexical entailments, but rather degrees of “logicalness”. He asks us to consider a list of implications, including:

$x$  is a bachelor  $\rightarrow x$  is unmarried

$x$  is a closed curve in the plane  $\rightarrow x$  has an area

$x$  is very *Adj*  $\rightarrow x$  is *Adj*

Necessarily  $P \rightarrow P$

$x$  is identical with  $y \rightarrow y$  is identical with  $x$

There is an  $F$  that  $V$ 's every  $G \rightarrow$  every  $G$  is  $V$ 'd by some  $F$

$P$  and  $Q \rightarrow P$

According to Lycan, the list displays an order of logicalness (p. 393). This leads him to abandon the idea of a small subclass of genuinely logical words within the English lexicon (p. 398). Thus, as Harman, assumptions regarding logical terms lead Lycan to a negative answer to the question of logic and natural language.

Finally, Glanzberg (n.d.) argues that the syntax and semantics of natural language do not yield a logical consequence relation, when the latter is understood to be as restrictive as standard first or second order logic. Glanzberg presents several arguments to that effect, one relying on the claim that natural language does not distinguish logical terms. Glanzberg acknowledges that his arguments might be less conclusive with respect to a more permissive view of logic. Yet, as Harman and Lycan, his main arguments pertain to the thesis that natural language has a relation of logical consequence akin to the one of traditional logical

---

<sup>2</sup>In other writings Harman does ascribe logical forms to natural language sentences, where the former are very close to standard first order logic. But the logic thus ascribed to natural language is chosen through meta-theoretical considerations, and is not determined by mere linguistic phenomena. See (Harman, 1972c, 1972b).

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

systems. Thus, *a fortiori*, logical terms play a dominant role in the considered notion of logical consequence.

To conclude this modest survey, I have distinguished between two projects that are relevant to the question of logic and natural language. What I have called the “traditional project”, exemplified in the work of Frege, Tarski and Carnap, logic is viewed as a methodology for science. As logic is meant to secure the inferences made, it is generally of restrictive nature: the variety of syntactic constructions is kept to a minimum, and the logical terms constitute a small limited class. The traditional project has given us standard first and second-order logics (as well as some alternative logics, such as intuitionistic logic, not discussed here). In the so-called “linguistic project”, on the other hand, the aim is to characterize natural language. This in essence is an empirical project concerned with linguistic phenomena in general. One way of investigating language is through its immediate or felt implications. In the linguistic project those are taken at face value and are aimed to be captured by formal means, and thus serve as the basis for a relation of logical consequence in natural language. While in the traditional project, natural language is handled with suspicion, in the linguistic project it is the subject matter of inquiry. In addition, while in the traditional project formal languages are the idiom in which secure inference is carried out, in the linguistic project they should be seen more as a mathematical tool in linguistic investigations. Thus, while the traditional project is wary of discussing logical consequence in natural language, the linguistic project assumes this can be done.

Lastly, we have seen some attempts to bring the two projects together, even if put in different terms than those I have been using. Harman, Lycan and Glanzberg have separately dealt with the question of whether any logic like the one (or ones) given to us in the traditional project can be found as a phenomenon in natural language. All three of them replied negatively. They all relied on the distinction between logical and nonlogical terms as necessary for logic, and denied such a distinction can be found in natural language. In general, what they look at are various features of conventional logic and show that natural

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

language does not exhibit them. Even though their perspective of the issue is very close to what I have in mind, and I agree with their bottom-line conclusions, their arguments are too limited for my purposes. This is not merely because I do not accept the distinction between logical and nonlogical terms as a necessary and distinctive feature of logic. It is because these studies do not get to the essence of the issues involved, and present merely contingent results. Change your choice of favorite logical system and your approach to natural language semantics, and the conclusions need to be reestablished and may no longer hold. In Section 10.1.3 I give a more detailed criticism of their main argument. Moreover, I propose to abstract away from specific features of logic and language, and ask what in the essence of natural language and of logic of the traditional project drives them apart. In the next section I will make an attempt at dealing with this question.

### 10.1.2 Making sense of the question: conceptually

Following the historical survey, I should now say what *I* mean by “logical consequence”, “natural language” and the former being in the latter. To these I add the notion of *formality*, which is also important for our purposes. I shall start by stating preliminary assumptions regarding each of these concepts, and in the next section make a connection with the philosophical perspectives I have described previously.

*Natural language.* In the philosophical literature I have been considering, the concept of natural language is commonly assumed without definition. There is usually no need for one, as there is in general agreement on the features of natural language relevant for the discussion. I shall assume, as do most of the writers I referred to, that natural language is part of the natural world, and is to be studied empirically. Further, it is usually agreed that English, French etc. are examples of natural languages.<sup>3</sup> In addition, for the purpose of the current discussion, I assume that a natural language is associated with a set of sentences, if

---

<sup>3</sup>Notwithstanding Chomsky (1975), who views such distinctions between languages as problematic.

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

not defined by it. We may treat natural language as an abstract grammar or as a cognitive capacity following Chomsky, but if the question of logical consequence in natural language stands a chance, we should have sentences to which we may ascribe various relations. The set of sentences will be usually given, however, through rules that are components of the grammar or theory of the language.

*Logical consequence.* First, we should indicate that logical consequence is a relation between sets of sentences and sentences.<sup>4</sup> The relation of logical consequence, I will assume, is determined by a system of rules, either semantic or syntactic. This makes plausible the treatment of logic as a tool or methodology. I will refer to the relation of logical consequence and the associated system of rules alternately. In previous chapters, I have accepted a conventional view of logical consequence as satisfying the conditions of *Necessity* and *Formality* (see Chapter 3, p. 3.2). For the purposes of the current discussion, I will concentrate more on *Formality*, which will be discussed below.

*Logical consequence in natural language.* Let  $S$  be the set of sentences that is or is associated with (some) natural language. Let  $R = \mathcal{P}(S) \times S$  (where  $\mathcal{P}(S)$  is the power-set of  $S$ ). The issue at hand is whether there is a non-trivial set  $LC \subseteq R$  that: (a) satisfies the requirements on logical consequence, and (b) is a phenomenon wholly determined by natural language. That is, there is a system of rules for  $LC$  that is part of the rules associated with the given natural language. Condition (a) is satisfied by some trivial subsets of  $R$ : the empty set, and arguably the minimal subset  $m$  of  $R$  satisfying reflexivity (for  $\Gamma \subseteq S$ ,  $\varphi \in S$ , if  $\varphi \in \Gamma$  then  $(\Gamma, \varphi) \in m$ ). To make the question interesting, I added the requirement that  $LC$  be non-trivial, where non-triviality will not be strictly defined. Condition (b) is

---

<sup>4</sup>For simplicity I accept the standard *relata* associated with logical consequence. Whether the premises form a set or a multiset, and whether there might be multiple conclusions does not bear on the issues discussed. As for taking sentences: perhaps the items involved should be more refined, involving propositions or contexts. For current purposes, we need to assume that sentences are involved—that logical consequence is a linguistic relation. Other kinds of entities might be involved, but this will not be assumed.

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

added to rule out relations of logical consequence that are artificially defined on sentences of natural language. For instance, one can presumably artificially assign forms to sentences of natural language, in a way that results with a relation satisfying the requirements of logical consequence.<sup>5</sup> There may be indeterminacy and theoretical considerations involved in the study of logical consequence, but if logical consequence is considered as a phenomenon in natural language it cannot be a mere theoretical artifact.

*Formality.* The condition of *Formality* states that logical consequence is determined by the forms of the sentences involved. In order to see whether such a condition can be fulfilled by a relation in natural language, we need some understanding of what forms of sentences may be. Now of course there is an accepted general characterization of forms of sentences: they are, or are represented by, formulas. These formulas belong to an formal language, which is part of an adequate formal system (the characterization of which is open for debate). A form is a form of some sentence if it captures its truth conditions, in a sense to be specified by the semantics used. Certainly these remarks are cursory, but there is no need at this point for further details of the formal language from which forms are taken or other components of the formal system. We may note that the conception of form depicted so far cuts across both the traditional and linguistic projects. As it stands, it bears no apparent conflict with the existence of logical consequence in natural language. What will put more weight on the question is looking into the motivations for *Formality*, as influenced by the traditional project.

### 10.1.3 Logic as methodology and natural language

In the rest of the discussion, I will endorse the traditional perspective on logic. I will see whether from this perspective, given the way the concepts involved were construed in

---

<sup>5</sup>The artificial assignment of forms to natural sentences can be a stage in a process of regimentation or formalization. Such processes lead to a relation of logical consequence in regimented or formalized natural language, as will be shown in Section 10.2.

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

the previous section, natural language has a logical consequence relation. Endorsing the traditional perspective does not constitute disagreement with the linguistic project in and of itself. Each project has its own aims and purposes. Recall that natural language is to be understood as a natural phenomenon to be dealt with by empirical methods. The linguistic project does exactly that: it investigates natural language with empirical means, reinforced by formal tools. Thus, the question I am interested in is whether the linguistic project can provide a logic acceptable by the traditional project: whether from the empirical study of language we can obtain a tool for correct reasoning, in accord with the conceptual specification above.

Now, on the one hand, it is indubitable that linguistic study reveals formal relations in natural language, in some sense of *formal*. Linguistic practice relies on the assumption that natural language, as any natural phenomenon, has an underlying structure that may be revealed. The formal methods of logic are appropriate for that kind of study. Thus, the linguistic orientation to logic suggests a way of maintaining that there is a logical consequence relation in natural language given the assumptions made so far concerning the concepts involved. There may be a relation in natural language which satisfies *Necessity* and *Formality*, if we don't add further requirements. The data in linguistic research includes the felt implications accepted by speakers. As *Necessity* is not the issue here, assume that the linguist can extract from the data a significant set of entailments that satisfy *Necessity*. Arguably, this can be done with adequately conducted research. A subset of the necessary entailments are then described in a formal system. The formal system includes a translation of natural language sentences into a formal language and a consequence relation that matches the given subset of entailments. As the translations can be considered to be forms of their natural language counterparts, *Formality* is thus satisfied.

Some of the writers mentioned in Section 10.1.1 (Harman, Lycan and Glanzberg) are reluctant to accept the result of such a process as a genuine natural language logical consequence relation. On their accounts, logical consequence is a relation more restrictive than

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

necessary entailments. And the process described does not require that there is such a distinction in natural language: any necessary entailment is a candidate to be formalized in the theory. If the formal language used to work out logical consequence in natural language is restricted to something in the vicinity of standard first or second order logic, many of the entailments will be left out—and so we have a relation which is more restrictive than necessary entailment. But those entailments that are not captured by the system are not distinguished from those that are by some feature found in natural language (this is argued extensively and in different ways by the three writers). The linguistic project, it is thus argued, does not present us with logical consequence in natural language.

Subsequently, I will propose a different line of argument towards a similar conclusion. Before that, I would like to place some objections to the HLG (Harman-Lycan-Glanzberg) argument.<sup>6</sup> The first objection has to do with assumptions on logical terms. Logical consequence is assumed in the argument to be a relation depending on a strict division of terms into logical and nonlogical. HLG then argue that there is no natural category of logical terms in natural language, and use this claim to argue that natural language does not have a logical consequence relation. In Part II of the dissertation I have opposed the assumption that logical consequence requires a strict division between logical and nonlogical terms. I proposed an alternative where a strict division is not assumed. I believe my approach is in line with the aims of the traditional project. Thus, it does not suffice to show that natural language does not have a category of logical terms to show that it does not have logical consequence. To be fair, some of the arguments given by HLG might be salvaged when excluding the strict division. Also, the strict division between logical and nonlogical terms is accepted in the conventional understanding of logic: the HLG arguments can be seen as aiming at the conventional conception of logical consequence. However, the thesis

---

<sup>6</sup>I should make clear that it is not just one argument that Harman, Lycan and Glanzberg present. Glanzberg alone offers three different arguments. What I am referring to is the general line of argument which is common to all three of them, and the main assumptions on which they rely.



## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

that natural language does not have a logical consequence relation becomes quite weak if it depends on an unnecessary assumption. I would like to suggest that if it is found out that natural language does not display a natural category of logical terms, it should raise the question of whether this category is absolutely necessary for our conception of logic rather than bring to the conclusion that natural language does not have a logical consequence relation.

The second objection comes from the side of natural language. In some recent studies in linguistics it has been proposed that language does indeed separate between logical and other entailments. Gajewski (2002) argues for a category of sentences that are L-analytic—true or false in virtue of form—as a special case of ungrammaticality, based on speakers’ intuitions. Presumably, his account can be extended to include entailments. Fox (2000, 2006) argues that the cognitive system contains a *deductive system* in which sentences are evaluated and ruled out if they can be proven to be contradictory. Fox’s characterization of the deductive system, as well as Gajewski’s characterization of the L-analytic sentences employ a distinction between logical and non-logical words, where logical words correspond roughly to the logical terms in standard first order logic. Thus, the contention of HLG that natural language does not have a natural category of logical terms may simply be false on the basis of these studies.

Finally, the main weakness of the HLG argument is that it is too superficial for my current purposes. I would like to know what in the essence of natural language, and what in the essence of logical consequence make it the case that there is or there isn’t logical consequence in natural language. The first two objections are related to this point. Even if objections can be made to the studies done by Gajewski and Fox, and they do not yield the desired category of logical terms, could it be the case that further studies identify a category of logical words? Should we say in that case that natural language does have a logical consequence relation? Or, could we conceive of a natural language different from English that exhibits such a relation? And, to repeat the first objection, is it essential

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

for a consequence relation to have a strict division of terms into logical and nonlogical? If not, could we then say that natural language does have a consequence relation? My claim, which I will argue for shortly, is stronger than the one substantiated by HLG: even when considering only the very fundamental aspects of logic and language, we come to the conclusion that natural language does not have a logical consequence relation.

Moving on to my main argument, I would like to assume the traditional perspective on logical consequence. This perspective brings with it certain principles, which I articulate below. I shall then consider two “candidates” for being logical consequence in natural language, provided by the linguistic outlook. I shall show that each of the candidates conflicts with at least one principle, and conclude that the linguistic project cannot provide a logic in natural language which is adequate for the purposes of the traditional project.

The notions of *necessity* and *formality*, which I used to describe logical consequence, are highly theoretic and philosophically loaded. I have previously adopted a possible worlds interpretation of the condition of *Necessity*. In addition, I have characterized *Formality* as a means to avoid metaphysical questions (Chapter 3). The idea was that logical consequence should be secure from contentious assumptions and stand on firm grounds. Here, I would like to take a step back, and look at the rudimentary principles that might motivate *Necessity* and *Formality* from the traditional perspective. I thus propose the principles of *practicality* and *safety*: one having to do with the aim of logic as a methodology, the second having to do with the quality of method and what the methodology is for.

1. *Logical consequence is practical*: logic is applicable in actual practices of reasoning.
2. *Logical consequence is safe*: logic is reliable, a logically valid argument is guaranteed to lead from truths only to truths.<sup>7</sup>

---

<sup>7</sup>Avron (1994) places two demands, *Faithfulness* and *Effectiveness*, on any formal system that is used to determine the validity of arguments. Paraphrased into the terms I have been using, they are:

- **Faithfulness.** If the formal system can be used to show that an argument is valid then this is actually

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

The first principle, the *principle of practicality*, will be objectionable by some. One may hold that as logic is concerned with ideal reasoning, there is no commitment to practicality (see (Wagner, 1987, p. 6)). But this defies the purpose logic has set itself, at least in the traditional project. A theory of logic that is impractical whatsoever cannot be viewed as a methodological discipline. This means, firstly, that logic needs to give us an explicit tool (such as a formal system) and not merely truths about what such a tool might be like (which rightfully belong to “metalogic”). Only when rules are given explicitly, can they serve as guidance.<sup>8</sup> The tool given might not be applicable for everyone or in every case,

---

the case.

- **Effectiveness.** If someone uses the formal system to show that an argument is valid then their success can be checked mechanically. Accepting Church’s thesis would mean that the set of arguments that can be shown to be valid by the formal system is an r.e. set.

(See pp. 229f.) My *principle of safety* is covered by Avron’s *Faithfulness*, given that logical consequence preserves truth. Avron’s *Effectiveness* is a reasonable addition, but is not needed for the current discussion. Avron also claims, separately from these principles, that a main feature of logical systems is that they are meant to be applied (p. 230), in line with my *principle of practicality*.

<sup>8</sup>The requirement that the rules be given explicitly has been made by members of the traditional project as Frege and Carnap. Frege, in the *Basic Laws*, demands “that all methods of inference employed be specified in advance” (Frege, 1964, p. 2). His reason for explicitness is not so much the practicality of logic, but, it seems, that the solidity of the mathematical edifice be apparent. Explicitness, for Frege, seems to be a mark of rigor and to carry with it epistemic value. The connection to the practicality of logic is not obvious, and I leave it open whether such exists for Frege (*cf.* (Coffa, 1991, p. 123)).

Carnap contended that “It is certainly possible to recognize from its form alone that a sentence is analytic; but only if the syntactical rules of the language are given.” (Carnap, 1937, p. 186). A possible interpretation of this quote yields the consequence that logic can be practical only if the rules of the language are given explicitly. This was a reply to Wittgenstein’s claim in the *Tractatus* that “It is the characteristic mark of logical propositions that one can perceive in the symbol alone that they are true; and this fact contains in itself the whole philosophy of logic.” (6.113). Carnap’s addition that the rules must be given comes from his position that there is a conventional aspect to form. Form is relative to a linguistic framework, so only given a framework can analyticity (or logicity) be determined. Carnap writes that the rules should be “given”, not “given explicitly”, but it is clear that explicitness is implied. The quote itself pertains to actual

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

but it does have to be directly connected to practice.

So logic is an instrument for reasoning. We reason all the time without using any tools, and at least some of the time we do it in a way that leads from truths to truths. The tools logic provides are meant to make sure that we reason correctly, that truths are guaranteed to lead to truths, and wherefore we have the *principle of safety*.

We can see how the principles of practicality and safety motivate *Necessity* and *Formality*. A relation or system of rules satisfying *Necessity* will *a fortiori* be safe. As for *Formality*: relying on just the forms of sentences suggests a practical way to evaluate arguments while maintaining necessity. Indeed, a formal system is the right kind of tool that can, if constructed in the right manner and given explicitly, be applied in particular cases of reasoning in a safe manner.

The question at hand is this: is there a relation in natural language that is in agreement with the principles of safety and practicality? Let us assume for the moment that there is such a thing as a relation of logical consequence in natural language. In such a case, as I have assumed earlier, this relation would be wholly determined by natural language: its discovery would not require methods over and above or different in nature from those used for studying natural language. As I assumed that natural language is an object of empirical inquiry, logical consequence in natural language should be amenable to empirical inquiry as well. Adopting a moderately realist viewpoint, we can distinguish between the theory of the phenomenon investigated and the phenomenon itself. This kind of distinction is made by Chomsky with respect to grammar. Chomsky indicates an ambiguity in the usage of the term “grammar”, that is:

the term ‘grammar’ is also used for the explicit theory, constructed by the lin-

---

practice of recognizing whether a sentence is analytic, so if the syntactic rules must be given, they must be given explicitly. There seems to be a closer connection between explicitness to practicality in Carnap’s case rather than Frege’s, although it is still only tacit.

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

guist, which purports to a theory of the rules and principles, the grammar in the first sense, that has been mastered by the person who knows the language. [...]  
The linguist's grammar is a theory, true or false, partial or incomplete, of the grammar of the speaker-hearer, the person who knows the language. The latter is the object of the linguist's study. He cannot, of course, observe it directly and can only attempt to construct a theory of the speaker-hearer's grammar, making use of whatever evidence he can obtain by observation or introspection, all such evidence, of course, being fallible and subject to correction and revision.  
([Chomsky, 1975](#), p. 303)

We can use Chomsky's distinction, substituting "logic in language" for "grammar". Let us adopt Chomsky's terminology, and refer to "logic in language in the first sense" and "the linguist's logic in language". The former is the purported phenomenon of logical consequence in natural language which is studied empirically by linguists. The latter is the outcome of the linguists' investigation, the theory they finally come up with, for logical consequence in natural language. Logic in language in the first sense and the linguist's logic in language are the main candidates for the sought-after relation that can serve as a tool for reasoning. However, in neither sense does "logic in language" fulfill both principles of practicality and safety:

- (a) Logic in language in the first sense does not satisfy practicality.
- (b) The linguist's logic in language does not satisfy safety.

Let us substantiate these claims in turn. (a) is a direct outcome of the distinction Chomsky made, applied to our case. Logic in language in the first sense is not directly observable by the linguist, or anyone else for that matter. Recall that in [Section 10.1.2](#) I assumed that the relation of logical consequence is determined or describable by rules. Indeed, a relation between sentences cannot be said to be applicable unless there is something like rules that

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

can be applied on given arguments. So suppose that there is a relation in natural language, describable by certain rules, that is “safe”: some fixed relation that preserves truth. The rules describing this relation are never explicit: our access to it is always through theory. So, after all, the phenomenon itself cannot serve as a tool, and is thus impractical.

The basis for (b) stems from the fact that linguistics is an empirical science, and is further strengthened by the nature of the science of linguistics. The theory given to us by the linguist might be practical enough: we can apply it to language use in particular cases. However, would we really want to use linguistic theory in our reasoning in mathematics and science, as required by the traditional project? Should we refer to recent studies in linguistics when claiming to have made a valid argument? Accepting a conclusion of some argument on the basis of current linguistic theory does not seem sufficiently safe. Current linguistic theory is bound to give its place to another at some point in the future. Should we want to make our reasoning in other sciences subject to revision pending on the advance in linguistics? Linguistic theory, as any empirical science, is in Chomsky’s words “fallible and subject to correction and revision”. In addition, linguistics is rather a “soft” science. Despite increasing rigor in method, it is hardly reliable from the point of view of mathematics and other “hard” sciences where we may conduct our reasoning. We may mention some of the impediments for attaining of a safe relation of logical consequence through linguistic theory: the reliance on statistical evidence, the high dependency of the interpretation of natural language sentences on context and the vagueness of some linguistic concepts such as acceptability and grammaticality.

The traditional project, it might be claimed, should not aspire to ideal, hermetic safety—especially by those proponents of the traditional project that have given up the concept of the *a priori*. From a Quinean naturalistic viewpoint, there is nothing over and above the certainty of the empirical sciences. Therefore, one might settle with what is given by linguistic theory. I think that such an approach is too much of a compromise for the traditional project. Linguistic theory might provide us with a tool for reasoning, but far

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

from the one logic was intended for. In its current state, it is hard to see how it can function as an instrument for reasoning in other sciences. It is also hard to see how a more advanced state of the linguistic discipline could provide a tool for reasoning. In the next section I will attempt to show that such a compromise is not necessary, and that an artificial logic imposed on natural language is an appropriate tool.

As noted earlier, Quine himself viewed logic, or formal systems in particular, as a means to facilitate theorizing. This view does not imply that logical consequence is a relation in natural language, at least not in the terms I have been using. Logic, in this view, is determined by theoretical considerations, such as simplicity and completeness, over and above empirical data. It is a theoretical construct rather than a phenomenon in natural language. Now it might be claimed that empirical theories are not completely determined by empirical data, and theoretical considerations are always involved. But still, in the case of Quine's view of logic, it seems that theoretical considerations in the case of logic do not merely function as arbiters between empirically equivalent theories in the case of indeterminacy, but they determine the basic features of logic (such as being first-order and extensional).

There is one more avenue yet to explore. Strawson (1969-70) maintained that there is room for research in *non-empirical linguistics*, where the "essential grammars" of various types of languages are constructed through *a priori* considerations alone. Perhaps much can be achieved in the same manner with respect to logic in natural language, and the safety of logic will not be impugned for use of empirical methods. So perhaps logical consequence is a relation in natural language which is distinguished from other natural language phenomena in that it can be after all discovered by safe means. For instance, if the conditions of *Necessity* and *Formality* were to already determine a relation for natural consequence for any given language without need of empirical investigation, linguistic theory could after all provide us with a safe tool for reasoning.

Unfortunately, it appears that the only a trivial relation of consequence can be achieved

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

in this way. The attribution of forms to natural language sentences must be an empirical matter, if logic is considered a part of natural language, and natural language part of the natural world. Now, as Strawson suggests, we can make progress by considering *a priori* the the various logics of different language types. Nonetheless, logic can be a tool for reasoning only if applied to a concrete language. Characterizing the type of a specific language, so that its logic can be brought to light, will have to be done through empirical study. Only trivial cases, where no use of specific sentences' forms is needed to determine validity, can be accounted for non-empirically. The only such cases are when the conclusion is one of the premises.

I conclude that neither logic in language in the first sense nor the linguist's logic in language live up to the standards of the traditional conception of logic. As the study of language as an empirical subject provides us no other alternative, there cannot be a relation of logical consequence that is a part of and determined by natural language.

### 10.1.4 Conclusion

The methods of study of natural language developed in the second half of the 20th century, which employ formal systems of logic, make it natural and convenient to speak of the "logic of natural language". On the other hand, logic in the traditional project requires features that cannot be found in natural language. Only superficial features are common to logic in the traditional and linguistic project, while a deeper philosophical outlook drives them apart.

Coffa (1991) describes the semantic tradition from Kant to Carnap as concerned with finding a way to save the *a priori* without recourse to Kantian intuitions. One of the benefits of saving the *a priori* is that logic can maintain its special, solid status in science. In the latter part of the 20th century, however, largely due to Quine's influence, the *a priori* began to lose its appeal. Logic's special status, if it can be maintained, needs to find a



## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

different source of support. In this section I have shown that this source of support cannot be linguistic study.

Quine himself, I have claimed, treated logic as a theoretical construct rather than a natural phenomenon. This claim presupposes that within a theory, a distinction can be made between those elements that pertain to the phenomenon investigated and those that are constructs or artifacts. I admit that this distinction needs grounding, which I cannot give here. However, I do contend that there is a difference between what is set by us in a theory and what is given by nature, even if only a vague one.

The next section will offer a more positive outlook on logic and natural language. Natural language can have a logic, but only once it is formalized. Logic is thus conceived as an artificial and not natural part of language, and it is fixed through explicit commitments made by reasoners.

### 10.2 Logic as a system of commitments

Logical consequence, as I've claimed in the previous section, is not a natural phenomenon to be discovered in the empirical study of language. This leaves us with the alternative that logical consequence is artificial. Formal language, defined by explicitly stated rules, resists the shortcomings of natural language, as it is both safe and practical. Practicality comes from having explicitly stated rules which can be applied to determine validity – at least for some arguments if not for all. Those same rules are as safe as can be, since they define the language. Application of the rules, we can assume, is safe enough (we can always make mistakes, but in simple cases in standard formal systems we can assume we usually don't, especially if we consider a community of thinkers).

Should we give up reasoning in natural language, and restrict ourselves to formal languages? Indeed, this is one of the upshots of the Fregean project mentioned in the previous section. However appealing this suggestion might sound to logicians, it is doubtful that

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

natural language can be given up easily. Shapiro points out that no one has in fact used a formal language for extensive science or philosophy, at least not all by itself (Shapiro, 1998, p. 136). The reasons for this may be varied, and it is unclear whether this is a contingent or a necessary fact about human reasoning.

Putting aside the normative question, of whether we should reason in natural language, the view I presented raises a descriptive question. Even in the most rigorous scientific or mathematical disciplines, a formal system is not usually pronounced at the outset. If logical consequence is not in natural language, and it is created only when explicit formalization is carried out, it might be worrisome if we need to dismiss as “non-logical” almost all reasoning as it is actually practiced. My response to this worry is that I propose an ideal of correct reasoning that need not always be fully attained. In some cases, there are accepted rules or constraints, that remain, however, implicit for the most part. In other cases, rules about the use of some words are specified, but not as a complete formal system, leaving other rules implicit. Mathematics is a discipline where, even though the ideal I describe is not practiced, there are both explicit and implicit rules or constraints. Definitions are examples of the former: they are invariably given explicitly. As for the latter, we can consider the semantics of conditionals, especially in the scope of a universal quantifier. It is well accepted in mathematical proofs (even if only implicitly) that *Every  $S$  is  $P$* , paraphrased as *For every  $x$ , if  $Sx$  then  $Px$* , is true in the case where there are no  $S$ s—this is an example where an ordinary language term (the conditional, or the *if...then* locution) is rigidified and has, so to speak, semi-formal use.

Moreover, even in the ideal I propose, the shift to formal languages need not be a drastic one. As I will suggest later on, natural language itself can serve as the basis for a formal language through the process of formalization, which consists in committing to fixed constraints on language in subsequent reasoning.

Various writers have discussed criteria for adequate formalization, i.e. translation of natural language texts into logical notation. They differ with respect to whether: (a) these

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

criteria should be semantic (i.e. truth-conditional) or inferential. (Sainsbury, 1993; Baumgartner & Lampert, 2008) promote the former and (Peregrin & Svoboda, 2012) promote the latter; (b) reflective equilibrium is involved in the process: (Resnik, 1985; Peregrin & Svoboda, 2012) are examples promoting reflective equilibrium, and (Thagard, 1982; Baumgartner & Lampert, 2008) are in opposition; and (c) whether, in general, findings from empirical psychology should have substantial impact on logical practice: (Thagard, 1982) is for, and (Resnik, 1985) is against. I will not get into these issues here, but rather lay out, in very general lines, a certain attitude towards formalization and its place in reasoning.

Thus, in this short section I would like to lay the basis for an approach that combines the idea of semantic constraints presented in Part II with a Carnapian outlook. Typically, formalization consists in translation to a formal language, where the latter has a fixed vocabulary divided into syntactic and semantic categories, and a distinguished subset of the vocabulary consisting of logical terms: the fixed terms of the system. But from a more general viewpoint, what we do when we formalize is use rules or postulates to fix or rigidify certain aspects of language. These rules and postulates, I would like to suggest, are commitments we take up with respect to language in argumentation and discourse.

In Part II I have proposed a framework for formal languages which is based on the notion of *semantic constraints*. A semantic constraint restricts the possible interpretations of terms in the language without necessarily fixing them completely. Among the semantic constraints we may have restriction on grammatical category, or items giving a fixed interpretation to some terms, but we may also have constraints fixing the mutual dependencies between terms (such as “the extensions of *red* and *green* are mutually exclusive”). Semantic constraints of all sorts can be viewed as commitments of those partaking in discourse or argumentation.

The more commitments we make, and the stronger they are, the stronger the consequence relation received. A language can be very rich in terms of the variety of sentences it has, but only when commitments are made do logical validities arise. Semantic constraints allow us to be very subtle about our commitments: some aspects of the meaning of terms

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

can be committed to without commitment to others—terms need not be completely fixed for their meaning to be constrained.

Commitments have a conventional aspect: they are a product of an act which involves some choice. No set of commitments is necessary. A system of semantic constraints, in this sense, functions as a Carnapian linguistic framework. Once a commitment to a semantic constraint is made, there is no question of its truth: it functions as part of what defines the linguistic framework chosen to be employed. The semantic constraints committed to are not completely arbitrary, and may contain terms for which there was a prior use that is to be retained, at least to some extent. Once a certain term appears in a constraint of the system, some aspect of its meaning is fixed by the constraint and can be relied on in logical reasoning. The constraint does not need to accurately capture the perceived meaning of the term, and for this reason we can view the constrained term as *rigidifying* and even *replacing* the non-constrained one in our reasoning. On the other hand, semantic constraints may coincide with an empirically confirmed theory of language. However, their status as giving the form of the language for logical purposes would stem from being committed to, and not from empirical justification.

The ideas presented here, of an explicitly stated linguistic framework, its conventionality and its epistemic status are all Carnapian.<sup>9</sup> The idea of semantic constraints is a point of departure from Carnap's works, but still, I think, very much in his spirit. Carnap's systems rely on a distinction between logical and descriptive terms in the definition of logical validity. The wider notion of analyticity is explicated by Carnap through the use of meaning postulates, that are similar in their function to semantic constraints. Carnap's notion of analyticity is thus parallel to my notion of logical consequence (see Part II, Chapter 6). I should note that Carnap's treatment of the notion of *formality* is different. While Carnap views only his narrow concept of logical consequence, and not analyticity, as formal, in my

---

<sup>9</sup>See e.g. (Carnap, 1937, pp. xiv-xv) for the conventional aspect of logic. I rely on Creath's interpretation of Carnap, as in (Creath, 1992).

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

own view form is determined by semantic constraints (rather than by the logical vocabulary), and so I view my wider concept of logical consequence as formal. This, I think, is the only place where there is a significant divergence of my view from Carnap's.

Semantic constraints may be chosen according to their utility in scientific and other discourse—this, again, is the same in Carnap. Carnap describes a process of explication which results in replacing a commonly used term with another one, bounded by the rules of the system. Carnap requires the following of an adequate explicatum for a given explicandum: (1) similarity to the explicandum, (2) exactness, (3) fruitfulness, and (4) simplicity (Carnap, 1962, pp. 5-8).

Requirement (1) is fairly obvious, given the nature of explication. Carnap does not require identity between the meanings of the explicatum and the explicandum, as the explicatum is an exact concept meant to replace a usually vague one (p. 5). Carnap explains requirement (2) as follows: “The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an *exact* form, so as to introduce that explicatum into a well-connected system of scientific concepts” (p. 7). That is, explication is a process of formalization, in the sense that it takes expressions from natural language and imposes on them explicitly stated rules. Requirement (3) is prompted by Carnap's pragmatic approach towards linguistic frameworks. Finally, Carnap adds that requirement (4) is of lower priority than (1), (2) and (3). The simplicity of a concept, he explains, is measured by the “simplicity of the form of its definition” and by the “simplicity of the forms of the laws connecting it with other concepts” (*ibid*).

Carnap's requirements for adequate explication can be adopted when setting up a framework of semantic constraints. Work that has been done in Part II, Chapter 8 can be used in giving more substance to the notions of similarity and simplicity from requirements (1) and (4). In the aforementioned chapter I have discussed the issue of fixing or constraining a term in a manner faithful to its intended meaning, in a model-theoretic framework. Even though we might not want to capture the exact meaning of the explicandum, there can be

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

an intended simplified meaning which we try to capture with the explicatum. So, the similarity of the explicatum and the explicandum can be achieved through fixing or constraining the explicatum in a manner faithful to the simplified intended meaning. Further, I have discussed in the same chapter conditions on semantic constraints that may be employed so as to keep the system simple and devoid of excessive metaphysical assumptions.

It may be instructive to compare Carnap's idea of explication with Quine's idea of regimentation. Quine proposes a process of regimenting natural language to make it more apt for logical purposes. This process consists in paraphrasing sentences in ways that reduce ambiguities while keeping the meaning intended by the utterer of the sentence. The paraphrase need not be synonymous with the original sentence, as it will typically be less ambiguous. The paraphrase is then transformed into logical notation of a canonical idiom, which for Quine would be the language of standard first order logic (Quine, 1960, pp. 158ff).

Both Carnap and Quine describe a process which includes simplification and does not result in synonymy, or at least does not aim to do so. Carnap's explication typically works at the level of terms (words)—usually so-called “descriptive” terms used in science such as *temperaure*, and *probability*. Quine's regimentation works at the level of sentences, and mainly concerns grammatical structure and logical terminology.

Another difference between the two approaches, more relevant to my current concerns, has to do with the nature of the end-result. For Quine, the paraphrase is still part of ordinary or semi-ordinary language, while for Carnap the result of explication belongs to a constructed system of scientific concepts. For Quine, the result of regimentation is also associated with rules regarding logical notation. But as the rules are themselves stated in ordinary language, he does not perceive the regimentation as transcending ordinary language. The difference between the approaches may seem to be insubstantial at first, as the rules Carnap alludes to are also formed, in the end, in ordinary language. But the difference in emphasis is telling. While Quine stresses the affinity between the end-result and ordinary language, and the difference is merely quantitative (less ambiguity in the

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

end-result), with Carnap it appears that the transition to a systematized language is of much more significance. This divergence in attitudes reflects Carnap and Quine's different approaches to the status of the rules of a formal system: for Carnap they are constitutive of analyticity—they define a new language, while for Quine they have no such privileged role.

Thus, I adopt from Carnap the importance given to an explicitly defined formal system. Nonetheless, there is still a place for natural language, as the basis on which a formal system can be constructed. Logical consequence, by the position I am advocating, is imposed on natural language through a process of formalization, which consists in explicit commitments made with respect to language, expressed in the form of semantic constraints. The commitment to a linguistic framework precedes (in principle if not in practice) all argumentation and discourse. In this sense, the validities of the system can be said to be *a priori*. This status, however, is not something to be intuited or discovered, but is set out by decision.

Quine has criticized the explanation of analyticity through the notion of a semantical rule:

Semantical rules are distinguishable, apparently, only by the fact of appearing on a page under the heading 'Semantical Rules'; and this heading is itself then meaningless. (Quine, 1951, p. 33)

For Quine, merely being distinguished by the system as a semantical rule carries with it no explanatory force. Quine's criticism would stand were we out to describe a preexisting distinction in language. Indeed, one might think that with a list of semantic constraints (or 'semantical rules') we are aiming to capture some underlying features of a given term which are constitutive of its meaning, and in that way surface an underlying distinction between the analytic and the synthetic. This is not the view I am advocating. Instead, when spelling out the semantical rules we are *forming* a distinction. We are separating what we are willing to commit to at the outset of anticipated discussion from what we are not committing to. Semantical rules, in this perspective, are distinguished by the fact that

they are committed to, and this commitment is carried out precisely by putting them on a list with the appropriate heading.<sup>10</sup>

Unfortunately, further consideration of the epistemic status of semantic constraints will take us too far afield, and would require examination of the relation between implicit definition, stipulation and the *a priori*. This subject is being dealt with by contemporary writers (e.g. in the fairly recent book (Boghossian & Peacocke, 2000)). I leave the discussion of the epistemic status of semantic constraints in light of contemporary research for future work.

### 10.3 Revisiting logic-as-model

In Part I, Chapter 3, I adopted Shapiro’s *logic-as-model* approach ((Shapiro, 1998, 2006), see also (Corcoran, 1974; Cook, 2002)). According to this approach, a formal system (including a formal language, deductive system and model-theoretic semantics) is a mathematical model of correct reasoning in natural language. Specifically, formal languages model natural language, and deductive systems and model-theoretic semantics model various aspects of natural language that have to do with correct reasoning. Up until now I spoke about *logical consequence* rather than *correct reasoning*. I assume that the former is a theoretical notion intended to capture the latter. Thus, I assume that the conditions of *Necessity* and *Formality* are desirable features of an appropriate theoretical notion of correct reasoning (see Chapter 3).

In Section 10.1.3 I denied that logical consequence as a *formal* relation is a relation in (unregimented) natural language. Only in the context of an explicitly fixed formal system

---

<sup>10</sup>Quine leaves room for “the explicitly conventional introduction of novel notations for purposes of sheer abbreviation” where synonymy (and thus analyticity) is intelligible, while disputing the intelligibility of other kinds of synonymy (Quine, 1951, p. 26). I am in agreement with Quine insofar as I am not claiming there is any other source of analyticity. At the same time, I do not confine the “explicitly conventional introduction of novel notations” to just “purposes of sheer abbreviation”.



## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

or set of semantic constraints can we speak of logically valid arguments. We can attribute logical consequence to natural language by explicitly stating for it semantic constraints, but we then *transform* natural language. We perform an act of rigidifying and regimentation. In this picture, formal systems are still conceived of as tools for reasoning, yet they are not merely descriptive tools. Logical consequence can thus be said to be related to natural language only *indirectly*, always relative to a regimentation.

My view of formal systems is in tension with logic-as-model: if there is no logical consequence in natural language, then how can formal systems model it? In fact, my view is in tension with a whole range of well-received theories on the relation between formal language and natural language. The common ground of the views in this range is the contention that formal systems in some ways capture logical consequence in natural language, and by which elements in formal language represent features of natural language. In what follows I discuss this tension, and explain how some aspects of the logic-as-model approach can still be maintained even when formal systems are not taken to model logical consequence in natural language.

Roy Cook describes a continuum of views on the connection between formalization and the actual everyday discourse ((Cook, 2002), pp. 233-234). Cook presents this idea in the context of the formalization of vagueness, but it may also be formulated generally. All views share the contention that the formal system is a rigorous and precise tool for the investigation of validity in natural language. They are ordered by the proximity they find between natural language and formal language. On one end there is the traditional view of logic as description, according to which the formal system describes what goes on in natural language. The more accurate the description, the better the formalism. On the other end of the line lies the instrumentalist view of logic, according to which the formalism mirrors little or nothing going on in natural language. It merely tells us which arguments are valid. Somewhere in the middle there is the view of logic as model, advocated by Shapiro (Shapiro, 1998, 2006) and further developed by Cook. By this view, the formal system

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

serves as a mathematical model for natural language. Some of the features of the model, the *representors*, correspond to features of what is modeled, and others, the *artefacts*, do not and merely serve as ingredients of the model that make it simple and coherent. The formalism is not evaluated by its accurateness, but rather, there is a balance between simplicity and closeness of fit ((Shapiro, 1998) p. 138). The model will only be adequate, however, if it predicts the right results concerning the phenomenon modeled through correct representation of the relevant features.

As an example, we may look at model-theoretic semantics as part of the formal system, and take models to represent possible worlds or something of the like (see Chapter 3). We may then take the elements of models to represent objects in possible worlds. If we take mathematical objects such as sets as the elements of models, that would be an artefact of the system: letting sets represent cats, street signs, electrons and all other objects makes the model simpler and easier for mathematical handling.

The view of logic as model appears to be superior to those at the extreme ends of the spectrum. It relieves the pressure of the descriptive view to account for every detail in natural language, and yet purports to give more insight into the workings of logical consequence than the instrumentalist approach does.

Now, if there is no logical consequence in natural language, as was claimed in Section 10.1.3, then formal systems cannot model it. Any rules determining a logical consequence relation would not be a phenomenon in natural language. I have suggested that rules for a logic should be in the form of explicitly expressed commitments. Specifically, the boundary between logical and nonlogical terms, or the choice of semantic constraints if the framework of Part II is accepted, are not to be found in natural language. Thus, as elements of a formal system, they cannot be representors. But neither can they be artefacts, if logical consequence is to be considered a formal relation. On the logic-as-model approach, the extension of logical consequence in the formal system would reasonably be supposed to represent the extension of logical consequence in natural language. If logical terms or

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

semantic constraints are an artefact, we have an artefact as a primary parameter deciding the extension of a representor: different choices of logical terms or semantic constraints lead to different extensions of logical consequence. It then turns out that one of the elements *determining the extension* of logical consequence in the formal language has no grip on the phenomenon modeled. Other artefacts do not have the same effect on the outcome of modeling. The identity of other artefacts, such as the elements of models, does not make a difference to the extension of logical consequence: take elements of models to be numbers, sets, tables and chairs - what matters is how the interpretation function works. However, it makes a difference whether you take the quantifier “there are infinitely many” (or anything equivalent) as a logical term — the validity of some arguments would rise or fall according to that choice.

Shapiro explains that logic-as-model does not need to assume a realist position according to which there is one correct underlying logic to natural language. Nor does it need to assume a purely descriptivist stance towards logic: formal systems can also be seen to model what *ought* to be logical consequence in natural language. So logic-as-model is not excluded on the basis of there being no one correct underlying logic to natural language. Recall also that in Section 10.1 I referred to a very specific understanding of the notion of logical consequence in natural language: it was understood to be a natural phenomenon describable by the tools used to investigate natural language. But logic-as-model could work using a different notion of logical consequence in natural language, for instance, one that is understood to be a result of artificial regimentation or of taking up commitments, as described in Section 10.2. So perhaps logical terms, or semantic constraints, can be seen after all as representors, when what is modeled is the outcome of regimentation.

Let us take an example and see how such modeling could work. Take an argument in natural language, say the following:

All Greeks are mortal

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

Socrates is Greek

So, Socrates is mortal.

On the basis of the claims made in Sections 10.1 and 10.2, this argument in itself is neither logical valid nor logically invalid. However, imposing some logical consequence relation on English, the question of validity will be determined. Then we can model this logical consequence relation using a formal system: the argument will be translatable to a formal language, and the validity of the translation will correspond to the validity of the original argument in English supplemented with a logical consequence relation.

The problem with this procedure is that the imposition of a logical consequence relation on natural language in the manner described in the previous section already includes the act of formalization, that is, a formal system is already employed. The process of formalization includes setting rules to the language and translation into formal notation. After being formalized, the argument above standardly becomes:

$$\forall x(Ax \rightarrow Bx)$$
$$Ac$$

---

$$Bc$$

with the assumption of the relevant rules of interpretation for the sentences appearing in the argument. Now, when modeled by a formal system, it would stay just the same, as it is already in a mathematical form, simplified and idealized enough to be part of a model. However, the process of formalization is not that of modeling a logical consequence relation, but rather that of forming one. And there is no need for modeling the result in a formal system, as the result is already in one.

Furthermore, note that there are common guidelines to modeling and formalization. If we look at Carnap's requirements for adequate explication (which we employ in the process of formalization, see Section 10.2 and (Carnap, 1962, pp. 5-8).), we can see that

## CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE

they also basic to the approach of logic-as-model: (1) *Similarity*: logic-as-model requires similarity between the model and what is modeled, exhibited in the elements of the model that are *representors*. (2) *Exactness*: A model is constructed in an exact manner so it is easier to handle. (3) *Fruitfulness*: A model is constructed for the purpose of investigating some phenomenon, and should be fruitful at that. (4) *Simplicity*: A simple model has an advantage over a complicated one. Again, modeling after formalization has been carried out seems to be superfluous.

The logic-as-model approach and formalization as I have described it come apart in their relation to logical consequence. While the logic-as-model approach models it, formalization, on my approach, creates it. Yet we can adopt some important insights from the logic-as-model approach. The logic-as-model approach can be fully endorsed when it comes to the models used in model-theoretic semantics. In previous chapters, I adopted the view that models represent possible worlds in some way. Indeed, the view of models as representing possible worlds is of crucial importance if *Necessity* is to be captured. The distinction between representors and artefacts is useful in this context: it is important to distinguish between those features of a model that represent a feature of a possible world (possibly, its size) and those that do not (say, the specific properties of the members of the domain). Nonetheless, when it comes to the formal language it is harder to distinguish between representors and artefacts. Assuming that when formalizing we commit to a new use of language, a strict separation between the parts that describe the old use and those that are merely artefacts is inappropriate.

I conclude that there is tension between the logic-as-model approach and the view of formalization presented in previous sections. Still, logic-as-model can be employed with respect to some parts of a formal system, e.g. model-theoretic semantics. Indeed, I have relied on the logic-as-model approach in previous chapters (especially Chapters 3, 8 and 9): but in all cases the models of model-theoretic semantics were the issue. Viewing models as representing possible worlds under reinterpretations of the language is consistent with my

## *CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE*

proposed view of formalization.

*CHAPTER 10. LOGICAL CONSEQUENCE AND NATURAL LANGUAGE*

## Chapter 11

# Conclusion

Logically valid arguments, it is sometimes claimed, are uninformative: since the conclusion is necessitated by the premises, is *contained* in them, no new information is added. Logic, however, is not just about evaluating arguments. It is first and foremost concerned with constructing systems in which arguments can be evaluated. Moreover, the choice of the rules of the system is not a trivial matter. I have suggested that a logical system is constructed and put into use through commitments made with respect to a language. Such commitments pertain to the meanings of the terms in the language, but might also include metaphysical assumptions to some extent. Commitments are made in the context of a certain discourse, and are best chosen in a way that facilitates reasoning in that discourse.

The commitments made, whatever their nature, separate between what is given as part of the framework for discussion and what is left open by the framework. Indeed, once a framework is decided on, a logically valid argument is in some sense uninformative, but this in no way implies that logic is irrelevant for the advancement of knowledge.

No doubt, there is a distance between ideology and practice. The systems logicians have offered, even those that make an impact, are not ubiquitous in scientific discourse. Moreover, reasoners in various disciplines and contexts are not likely to articulate their



commitments to a logical framework, in the manner I have proposed. Yet logical theory presents an ideal that can be attained in various degrees, and a conceptual apparatus to the disposal of rigorous reasoners and which improves understanding reasoning in general.

This work generalizes the conventional notion of a logical system. A logical system need not be in the form of a compendium of rules that rely on a distinction between logical and nonlogical terms. A logical system, more generally, sets constraints on the possible use and interpretation of a language. Since semantic constraints are phrased as sentences in the metalanguage, they can easily be seen as commitments: as claims about the object-language committed to by participants of a discourse.

The framework of semantic constraints, presented here, is still in its rudimentary stages. There is much further work to be done:

- First, there is the project of applying the framework in a useful manner to philosophical or scientific debates. The examples of systems of constraints I have brought were intentionally simplistic, so as to direct attention to the formalism. So there is place for the construction of systems that pertain to “real” discourse.
- A related project is that of translating existing logical frameworks into semantic constraints. I have shown how this can be done with standard propositional, first-order and modal logic. I have claimed that the framework of semantic constraints is a natural home for modal logic. Further research could show the same with respect to other systems dealt with by logicians.
- Another prospective project is the development of a parallel framework of semantic constraints in proof-theoretic semantics. A first step in this direction was taken in Chapter 7. However, there is much more conceptual work to be done.
- There is also the project of applying the new framework to linguistic theory. In Chapter 10.1 I distinguished between logic as providing a method for reasoning and

## CHAPTER 11. CONCLUSION

logic as providing a tool for the investigation of natural language. In this work I have been mainly concerned with the former. However, the machinery developed in Chapter 7 can be employed in linguistic studies. The concepts of *dependency*, *determinacy* and *category* could be especially useful in such studies.

Other avenues of further research involve undeveloped issues and assumptions occurring in this work. So, for instance, the condition of *Necessity* on logical consequence (p. 20) was assumed without qualifications. Indeed, this condition is uncontroversial—but there is much room for a critical evaluation of it. There is clear motivation for the requirement that logically valid arguments preserve truth: that truths do not lead to falsehoods. What is the motivation for requiring that truth-preservation be *necessary*?

Another unresolved issue concerns the view of logic as a system of commitments. This view was presented at a basic level, and there is work to be done regarding the epistemic status of such commitments. Specifically, as I have indicated in Chapter 10.1, further work would explore the relation of logical commitments in the form of semantic constraints to implicit definitions, stipulation and the a priori.

## *CHAPTER 11. CONCLUSION*

# References

- Avron, A. (1994). What is a logical system? In D. Gabbay (Ed.), *What is a logical system?* (Pp. 217–238). New York: Oxford University Press.
- Azzouni, J. (2003). The strengthened liar, the expressive strength of natural languages, and regimentation. *The Philosophical Forum*, 34(3-4), 329–350.
- Azzouni, J. (2006). The inconsistency of natural languages: how we live with it. *Inquiry*, 50(6), 590–605.
- Azzouni, J. (2012). Inconsistency in natural language. *Synthese*.
- Bar-Elli, G. (2004). Introduction. In G. Bar-Elli (Ed.), *Begriffsschrift (Hebrew translation)*. Jerusalem: Shalem Press.
- Barwise, J. (1969). Infinitary logic and admissible sets. *The Journal of Symbolic Logic*, 34(2), 226–252.
- Baumgartner, M., & Lampert, T. (2008). Adequate formalization. *Synthese*, 164, 93–115.
- Blackburn, P., de Rijke, M., & Venema, Y. (2002). *Modal logic*. Cambridge: Cambridge University Press.
- Blakemore, D., & Carston, R. (2005). The pragmatics of sentential coordination with *and*. *Lingua*, 569-589, 45–72.
- Blanchette, P. A. (2000). Models and modality. *Synthese*, 124, 45–72.
- Boghossian, P. (1997). Analyticity. In B. Hale & C. Wright (Eds.), *A companion to the philosophy of language* (pp. 331–368). Oxford: Blackwell.

## REFERENCES

- Boghossian, P., & Peacocke, C. (Eds.). (2000). *New essays on the a priori*. Oxford: Calderon Press.
- Bolzano, B. (1973). *Theory of science* (J. Berg, Ed. & B. Terrell, Trans.). Dordrecht, Holland: D. Reidel Publishing Company.
- Bonnay, D. (2008). Logicality and invariance. *The Bulletin of Symbolic Logic*, 14(1), 29–68.
- Carnap, R. (1937). *The logical syntax of language*. London: Routledge and Kegan Paul.
- Carnap, R. (1947). *Meaning and necessity*. Chicago: The University of Chicago Press.
- Carnap, R. (1948). *Introduction to semantics*. Cambridge, MA: Harvard University Press.
- Carnap, R. (1962). *Logical foundations of probability* (2nd ed.). Chicago: The University of Chicago Press.
- Carnap, R. (1963a). E. W. Beth on constructed language systems. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 927–933). La Salle, IL: Open Court.
- Carnap, R. (1963b). W. V. Quine on logical truth. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 915–921). La Salle, IL: Open Court.
- Casanovas, E. (2007). Logical operations and invariance. *Journal of Philosophical Logic*, 36(1), 33–60.
- Chomsky, N. (1955). Logical syntax and semantics: their linguistic relevance. *Language*, 31(1), 36–45.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1975). Knowledge of language. In K. Gunderson (Ed.), *Minnesota studies in the philosophy of science, volume vii: language, mind and knowledge* (pp. 299–320). Minneapolis: University of Minnesota Press.
- Chomsky, N. (1980). *From rules to representations*. New York: Columbia University Press.
- Coffa, A. (1991). *The semantic tradition from Kant to Carnap*. Cambridge: Cambridge University Press.
- Cook, R. (2002). Vagueness and mathematical precision. *Mind*, 111, 225–247.
- Copi, I. M. (1961). *Introduction to logic* (2nd ed.). New York: The Macmillan Company.

## REFERENCES

- Corcoran, J. (1974). Aristotle's natural deduction system. In J. Corcoran (Ed.), *Ancient logic and its modern interpretations*. Dordrecht: Reidel.
- Creath, R. (1992). Carnap's conventionalism. *Synthese*, 93(1/2), 141–165.
- Curry, H. B. (1942). The inconsistency of certain formal logic. *The Journal of Symbolic Logic*, 7(3), 115–117.
- Davidson, D. (1967). Truth and meaning. *Synthese*, 17(3).
- Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: Oxford University Press.
- Davidson, D. (2008). Semantics for natural languages. In S. Nuccetelli & G. Seay (Eds.), *Philosophy of language: the central topics* (pp. 57–63). Lanham: Rowman and Littlefield Publishers.
- Dutilh Novaes, C. (2012). Reassessing logical hylomorphism and the demarcation of logical constants. *Synthese*, 185(3), 387–410.
- Enderton, H. (1972). *A mathematical introduction to logic*. New York: Academic Press.
- Etchemendy, J. (1983). The doctrine of logic as form. *Linguistics and Philosophy*, 6(3), 319–334.
- Etchemendy, J. (1990). *The concept of logical consequence*. Cambridge, MA: Harvard University Press.
- Etchemendy, J. (2006). Reflections on consequence. In D. Patterson (Ed.), *New essays on tarski and philosophy* (pp. 263–299). Oxford: Oxford University Press.
- Evans, G. (1976). Semantic structure and logical form. In G. Evans & J. McDowell (Eds.), *Truth and meaning: essays in semantics* (pp. 199–221). Oxford: Clarendon Press.
- Feferman, S. (1999). Logic, logics and logicism. *Notre Dame Journal of Formal Logic*, 40(1), 31–55.
- Fox, D. (2000). *Economy and semantic interpretation, linguistic inquiry monographs 35*. Cambridge, MA: MITWPL and MIT Press.

## REFERENCES

- Fox, D., & Hackl, M. (2006). The universal density of measurment. *Linguistics and Philosophy*, 29, 537–586.
- Frege, G. (1964). *The basic laws of arithmetic: exposition of the system* (M. Furth, Ed. & Trans.). Berkely: University of California Press.
- Frege, G. (1967). *Begriffsschrift*. In J. van Heijenoort (Ed.), *From Frege to Gödel* (pp. 1–82). Cambridge MA: Harvard University Press.
- Frege, G. (1979). *Posthumous writings*. Chicago: University of Chicago Press.
- Frege, G. (1984). Thoughts. In B. McGuinness (Ed.), *Collected papers on mathematics, logic and philosophy* (pp. 351–372). Oxford: Basil Blackwell.
- Gajewski, J. (2002). *On analyticity in natural language*. Manuscript.
- Gentzen, G. (1969). Investigations into logical deduction. In M. Szabo (Ed.), *The collected papers of gerhard gentzen* (pp. 68–131). Amsterdam-London: North-Holland Publishing Company.
- Glanzberg, M. (**nodate**). Logical consequence and natural language. In C. Caret & O. Hjortland (Eds.), *Foundations of logical consequence*. Oxford: Oxford University Press.
- Gómez-Torrente, M. (2002). The problem of logical constants. *The Bulletin of Symbolic Logic*, 8(1), 1–37.
- Grice, H. (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.
- Hacking, I. (1979). What is logic? *The Journal of Philosophy*, 76(6), 285–319.
- Hanson, W. H. (1997). The concept of logical consequence. *The Philosophical Review*, 106(3), 365–409.
- Hanson, W. H. (2002). The formal-structural view of logical consequence: a reply to Gila Sher. *The Philosophical Review*, 111(2), 243–258.
- Harman, G. (1972a). Deep structure as logical form. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 25–47). Dordrecht-Holland: D. Reidel Publishing Company.
- Harman, G. (1972b). Is modal logic logic? *Philosophia*, 2(1-2), 75–84.

## REFERENCES

- Harman, G. (1972c). Logical form. *Foundations of Language*, 9(1), 38–65.
- Harman, G. (1984). Logic and reasoning. *Synthese*, 60, 107–127.
- Hodes, H. (2004). On the sense and reference of a logical constant. *The Philosophical Quarterly*, 54(214), 134–165.
- Keenan, E., & Westerstahl, D. (1997). Generalized quantifiers in linguistics and logic. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 837–893). Amsterdam: Elsevier.
- Kuhn, S. (1981). Logical expressions, constants, and operator logic. *The Journal of Philosophy*, 78(9), 487–499.
- Lycan, W. (1984). *Logical form in natural language*. Cambridge, MA: The MIT Press.
- Lycan, W. (1989). Logical constants and the glory of truth conditional semantics. *Notre Dame Journal of Formal Logic*, 30(3), 390–400.
- MacFarlane, J. (2000). *What does it mean to say that logic is formal?* (PhD thesis, University of Pittsburgh).
- MacFarlane, J. (2009). Logical constants. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Fall 2009).
- McCarthy, T. (1981). The idea of a logical constant. *The Journal of Philosophy*, 78(9), 499–523.
- McCarthy, T. (1987). Modality, invariance and logical truth. *Journal of Philosophical Logic*, 16(4), 423–443.
- McCarthy, T. (1998). Logical constants. In E. Craig (Ed.), *Routledge encyclopedia of philosophy* (Vol. 5, pp. 775–781). London: Routledge.
- McGee, V. (1996). Logical operations. *Journal of Philosophical Logic*, 25, 567–580.
- Menzel, C. (1990). Actualism, ontological commitment, and possible worlds semantics. *Synthese*, 85(3), 355–389.
- Montague, R. (1974a). English as a formal language. In R. Thomason (Ed.), *Formal philosophy* (pp. 188–221). New Haven: Yale University Press.



## REFERENCES

- Montague, R. (1974b). The proper treatment of quantification in ordinary english. In R. Thomason (Ed.), *Formal philosophy* (pp. 247–270). New Haven: Yale University Press.
- Montague, R. (1974c). Universal grammar. In R. Thomason (Ed.), *Formal philosophy* (pp. 222–246). New Haven: Yale University Press.
- Peacocke, C. (1976). What is a logical constant? *The Journal of Philosophy*, 73(9), 221–240.
- Peacocke, C. (2004). Understanding logical constants: a realist’s account. In T. Baldwin & T. Smiley (Eds.), *Studies in the philosophy of logic and knowledge* (pp. 163–208). Oxford: Oxford University Press.
- Peregrin, J., & Svoboda, V. (2012). Infinitary logic and admissible sets. *Synthese*.
- Popper, K. (1947). New foundations for logic. *Mind*, 56(223), 193–235.
- Quine, W. (1951). Two dogmas of empiricism. *The Philosophical Review*, 60(1), 20–43.
- Quine, W. (1953). Mr. Strawson on logical theory. *Mind*, 57(248), 433–451.
- Quine, W. (1956). Quantifiers and propositional attitudes. *The Journal of Philosophy*, 53(5), 177–187.
- Quine, W. (1960). *Word and object*. Cambridge, MA: The M.I.T. Press.
- Quine, W. (1986). *Philosophy of logic* (2nd ed.). Cambridge, MA: Harvard University Press.
- Read, S. (1994). Formal and material consequence. *Journal of Philosophical Logic*, 23(3), 247–265.
- Resnik, M. D. (1985). Logic: normative or descriptive? the ethics of belief or a branch of psychology? *Philosophy of Science*, 52(2), 221–238.
- Russell, B. (1905). On denoting. *Mind*, 14, 479–493.
- Sagi, G. (2011). Sher and shapiro on logical terms. In M. Peliš & V. Punčochář (Eds.), *The logica yearbook 2010* (pp. 199–210). London: College Publications.
- Sagi, G. (forthcominga). Formality in logic: from logical terms to semantic constraints. *Logique et Analyse*.
- Sagi, G. (forthcomingb). Models and logical consequence. *Journal of Philosophical Logic*.
- Sainsbury, R. M. (1993). *Logical forms*. Oxford: Blackwell Publishers.

## REFERENCES

- Shapiro, S. (1991). *Foundations without foundationalism*. Oxford: Oxford University Press.
- Shapiro, S. (1998). Logical consequence: models and modality. In M. Schirn (Ed.), *The philosophy of mathematics today* (pp. 131–156). Oxford: Oxford University Press.
- Shapiro, S. (2002). Necessity, meaning, and rationality: the notion of logical consequence. In D. Jacquette (Ed.), *A companion to philosophical logic* (pp. 227–240). Oxford: Blackwell.
- Shapiro, S. (2005). Logical consequence, proof theory, and model theory. In S. Shapiro (Ed.), *The oxford handbook of philosophy of mathematics and logic* (pp. 651–670). Oxford: Oxford University Press.
- Shapiro, S. (2006). *Vagueness in context*. Oxford: Oxford University Press.
- Sher, G. (1991). *The bounds of logic: a generalized viewpoint*. Cambridge, MA: MIT Press.
- Sher, G. (1996). Did Tarski commit ‘Tarski’s fallacy’? *The Journal of Symbolic Logic*, 61(2), 653–686.
- Sher, G. (1999). Is logic a theory of the obvious? *European Review of Philosophy*, 4, 207–238.
- Sher, G. (2001). The formal-structural view of logical consequence. *The Philosophical Review*, 110(2), 241–261.
- Sher, G. (2002). Logical consequence: an epistemic outlook. *The Monist*, 85(4), 555–579.
- Sher, G. (2008). Tarski’s thesis. In D. Patterson (Ed.), *New essays on tarski and philosophy* (pp. 300–339). Oxford: Oxford University Press.
- Smiley, T. (1982). The presidential address: the schematic fallacy. In *Proceeding of the aristotelian society* (Vol. 83, pp. 1–17). Blackwell Publishing (1982-1983).
- Strawson, P. F. (1969). The presidential address: grammar and philosophy. In *Proceeding of the aristotelian society* (Vol. 70, pp. 1–20). Blackwell Publishing (1969-1970).
- Tarski, A. (1933). The concept of truth in formalized languages. In J. Corcoran (Ed.), *Logic, semantics, metamathematics* (pp. 152–278). Indianapolis: Hackett.

## REFERENCES

- Tarski, A. (1934). Some methodological investigations on the definability of concepts. In J. Corcoran (Ed.), *Logic, semantics, metamathematics* (pp. 296–319). Indianapolis: Hackett (1983).
- Tarski, A. (1936). On the concept of logical consequence. In J. Corcoran (Ed.), *Logic, semantics, metamathematics* (pp. 409–420). Indianapolis: Hackett (1983).
- Tarski, A. (1986). What are logical notions? *History and Philosophy of Logic*, 7, 143–154.
- Tarski, A. (2002). On the concept of following logically. *History and Philosophy of Logic*, 23, 155–196.
- Tarski, A., & Givant, S. (1987). *A formalization of set theory without variables*. USA: Colloquium publications (American Mathematical Society).
- Thagard, P. (1982). From the descriptive to the normative in psychology and logic. *Philosophy of Science*, 49(1), 24–42.
- Tharp, L. (1975). Which logic is the right logic? *Synthese*, 31, 1–21.
- van Benthem, J. (1989). Logical constants across varying types. *Notre Dame Journal of Formal Logic*, 30(3), 315–342.
- van Benthem, J., & Bonnay, D. (2008). Modal logic and invariance. *Journal of Applied Non-Classical Logics*, 18(2-3), 153–173.
- Varzi, A. (2002, 1). On logical relativity. *Philosophical Issues*, 12, 197–219.
- Wagner, S. J. (1987). The rationalistic conception of logic. *Notre Dame Journal of Formal Logic*, 28(1), 3–35.
- Warmbröd, K. (1999). Logical constants. *Mind*, 108(431), 503–538.
- Wehmeier, K. (2004). Wittgensteinian predicate logic. *Notre Dame Journal of Formal Logic*, 45(1), 1–11.
- Wehmeier, K. (2012). How to live without identity—and why. *Australasian Journal of Philosophy*, 90(4), 761–777.
- White, M., & Tarski, A. (1987). A philosophical letter of alfred tarski. *The Journal of Philosophy*, 84(1), 28–32.

## *REFERENCES*

Wittgenstein, L. (1922). *Tractatus logico-philosophicus* (C. Ogden, Trans.). London: Routledge and Kegan Paul.