Title
Variables in Logic and Natural Language

Author
Boon, Ian

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Variables in Logic and Natural Language

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Philosophy

by

Ian Gregory Boon

2022

ABSTRACT OF THE DISSERTATION

Variables in Logic and Natural Language

by

Ian Gregory Boon
Doctor of Philosophy in Philosophy
University of California, Los Angeles, 2022
Professor Samuel John Cumming, Chair

It is standardly believed that some occurrences of expressions designate singularly, while other occurrences of expressions designate plurally. For instance, the singular expression `the student` may be used on an occasion to talk about one particular student, while the plural expression `the students` may be used on an occasion to talk about several different students. But it is rarely appreciated that several occurrences of expressions may together designate in a manner that is both plural and structured. Hence, the expressions `the students` and `their essays` may be jointly used on an occasion to talk about some students and some essays in a way that relates each student with the sub-collection of the essays that they authored. In other words, several occurrences of expressions may together designate a system of objects that may involve a non-trivial relation among them.

The first part of this dissertation investigates the semantic role that variables play in first-order languages. Variables are the most notable type of expression that partakes in this form of structured plural designation. I present a view according to which variables are interpreted by *sui generis* intensional entities. More specifically, I claim that variables express special concepts for thinking about the objects that make up a domain of quantification. These

concepts differ from other more familiar intensional entities precisely because they provide ways of thinking about structurally dependent systems of objects.

The second part of this dissertation investigates the semantic role that pronouns play in natural languages. Pronouns are commonly believed to be the closest natural language analog to variables. I substantiate this view by arguing that the same intensional entities used to interpret variables should also be used to interpret pronouns. In particular, I claim that some occurrences of expressions are governed by use-conditions that require a speaker to have one or more appropriate concepts of this kind in mind as intended subjects of discourse. Those communicative intentions are then used to fix the content of suitable subsequent occurrences of pronouns. After arguing for this view, I conclude by briefly sketching how it may be compositionally implemented in a standard semantic architecture with a few modest adjustments.

The dissertation of Ian Gregory Boon is approved.

Joshua David Armstrong

Gabriel Jae Greenberg

David B Kaplan

Sean Walsh

Samuel John Cumming, Committee Chair

University of California, Los Angeles

2022

To Kim,

who is the reason I managed to write anything at all

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This dissertation would have been impossible without the guidance and good company of those I have had the privilege of working with at UCLA.

I would first like to thank my committee. I could not have asked for a better advisor than Sam Cumming. He saw value in my work long before I could, and he has helped improve this project in more ways than he probably knows. He has an almost preternatural ability to see where a nascent idea is headed long before it gets there. I cannot count the times I thought he misunderstood me, only to later discover that I had actually misunderstood myself and that Sam had me right all along. And of course, his constant encouragement and extreme patience with my sometimes painfully slow rate of progress were always greatly appreciated. Gabe Greenberg pushed me the hardest when I needed it most and knocked me out of some of my worst intellectual habits early in my philosophical career (at least temporarily). He also taught me the value of making a point quickly, clearly, and forcefully and letting it linger for exactly as long as it needs to. The breadth of Josh Armstrong's knowledge in fields unfamiliar to me is frankly staggering. He always presented me with perspectives on issues that were somewhat different from those I was naturally inclined to take, which served as invaluable counterpoints as I tried to sharpen and clarify my own positions. Sean Walsh was a relative latecomer, and I will always find the fact that he expressed almost immediate interest in me and my work upon his arrival at UCLA to be deeply humbling. He, more than anyone else, taught me that technical rigor is only valuable in its proper philosophical context. And finally, David Kaplan's warmth and enthusiasm for philosophy were constantly reinvigorating whenever I felt my own passion waning. The way he teaches philosophy is truly inspiring, and any good pedagogical habit I have, I owe to him. I will surely strive to emulate him for the rest of my career, though I will never come close to the genuine article.

This dissertation also benefited immensely from conversations with my close friends and fellow graduate students at UCLA, who are among the best philosophers I know. Bill

Kowalsky is one of the most perceptive and engaging people I have ever had the pleasure of squabbling with over things both deep and trivial. Milo Crimi was the best kind of friend I could hope to make when I first moved to Los Angeles, even though he also served as a constant disturbing reminder that the philosophy of language existed long before the twentieth century and that it has not changed much in the intervening years. Amber Kavka-Warren was always willing to provide the helpful perspective of an intelligent person whose linguistic intuitions had not yet been poisoned by too much theory. Christian De Leon was always willing to provide the helpful perspective of an intelligent person whose linguistic intuitions had been poisoned by too much theory. Gabe Dupre taught me more about philosophy than I really care to admit, and so I will leave it at that. Gabby Johnson was always willing to call me out when I was wrong both morally and intellectually, and I am all the better for it on both counts. Torsten Odland was a frequent sounding board who would always patiently listen to my ill-formed ramblings when he had better things to be doing. But above all else, I would like to thank Kim Johnston. No one had to suffer my company and confusion more during the production of this dissertation, and I am forever grateful that she never held it against me.

I would also like to thank the many other faculty members (in both philosophy and linguistics) who helped make my time at UCLA both more enjoyable and more productive: Tyler Burge, Adam Crager, Katie Elliott, Pamela Hieronymi, Andrew Hsu, Tim Hunter, A.J. Julius, Ed Keenan, Tony Martin, Calvin Normore, Terry Parsons, Michael Rescorla, Jessica Rett, Yael Sharvit, and Sheldon Smith. Special thanks go to Sherri Roush, who has been a fixture in my philosophical education for over a decade and first taught me the joy of symbolic logic. Finally, I would like to thank everyone else who somehow contributed to the eventual existence of this document in some small way: Rima Basu, Tristen Cardwell, Jenna Donohue, Andrew Flynn, Catherine Hochman, John Kardosh, Tanya Kostochka, Kevin Lande, Esther Nikbin, Nefeli Ralli, Megan Robb, Levi Roth, Ayana Samuel, Vaheh Shirvanian, Femi Táíwò, Paul Talma, Aaron West, and Ekin Zeytinoglu.

VITA

2012 B.A. Philosophy (*Highest Distinction*), UC Berkeley, Winter 2012

2016 M.A. Philosophy, UCLA, March 2016

2017 C.Phil Philosophy, UCLA, September 2017

INTRODUCTION

The chapters that make up this dissertation originated as parts of separate papers on ostensibly different topics, and so I would like to take this opportunity at the outset to say a little bit about the underlying theme that unites them all. Language use undeniably involves the production and interpretation of a staggering variety of physically detectable patterns. So at a very high level of abstraction, a *particular* language can be identified with a collection of some of those patterns. But what features set the possible languages apart from *mere* collections of this kind? This is one of the deep foundational questions that the philosophy of language seeks to answer. And thankfully, essentially everyone agrees to a few standard criteria that make that task a bit more manageable. Most importantly, it is normally assumed that the possible languages all exhibit an extremely high degree of structural regularity that distinguishes them from other less uniform collections. This thought is ordinarily captured by imposing a requirement that a language is generable from some finite lexical base and a finite collection of structure-building operations. In other words, the possible languages must at least abstractly admit of some finite *grammar*. Further requirements for possible languagehood can then be stated in terms of additional constraints on the shape those grammars may take.

Some familiar examples of these additional constraints are entirely syntactic, in the sense that they merely concern how a grammar is related to the expressions it generates. For instance, the operations that a grammar makes use of are normally assumed to be well-behaved, in the minimal sense that they are computable. This requirement is then often further refined to ensure that the parsing or recognition problems for possible languages are not just computable but tractably so. And from a less mathematical angle, the familiar work in the generative tradition of linguistic theory assumes that the grammars for humanly possible languages are subject to human-specific biological constraints. This approach assumes that the brain realizes a high-level description of a generating device akin to a grammar and

that the possible shapes those generating devices may take are thereby largely genetically predetermined. Discovering the exact nature of those constraints has long been one of the stated tasks of empirical linguistic inquiry.

However, these familiar syntactic constraints are not the only ways of proceeding. Semantic approaches start with the observation that no possible language is a *mere* collection of expressions because they also need to be *interpreted*. These approaches then consider how a grammar and the expressions it generates are further related to the reality they represent or the thoughts they express. A guiding intuition behind some of these approaches is that the basic patterns of syntactic combination used in the grammar must, in some sense, mirror important combinations in the world or in cognition. Classically speaking, the most prominent example of this line of thinking at work is the traditional emphasis placed on the representational format of singular predication. A language's ability to naturally and systematically express singular propositions is normally assumed to require a clear distinction in the lexicon between singular and general terms. In other words, the language must distinguish between a class of expressions that must apply to one and only one object and a class of expressions that may apply to many objects. The familiar practice of using an expression like Fa to mean the object the singular term a applies to is among the things subsumed by the general term F is a clear case in point. In this example, expression types are principally distinguished in terms of their semantic roles. The asymmetry between those roles then projects into an asymmetry both in the lexicon and in grammatical combination – the only basic well-formed expressions are ones where a predicable stands in an asymmetric relation to a predicand. The semantic significance of asymmetric grammatical composition can then be straightforwardly taken to signal singular predication, which is normally understood as reflecting some important aspect of the basic composition of atomic facts or singular thoughts.

According to these approaches, there are broadly speaking two major requirements on the legitimate grammars of languages capable of expressing a robust range of singular proposi-

tions. First, the lexical expressions of the grammar must all be independently interpretable. And second, the grammar's basic modes of syntactic composition must respect something like the subject/predicate asymmetry. But while this basic picture may initially seem like an attractive way of viewing the elementary organization of language, it is extremely unclear how viable this strategy really is. In particular, it does not seem very difficult to find apparent exceptions to this general pattern. In fact, the very same kind of languages that will systematically make use of expressions like Fa to express singular propositions will *also* use expressions like $\forall xFx$ to express general ones. And this formula looks like it in part involves a grammatical combination Fx, with predicable F and predicand x. But here, the variable x does not appear to be independently interpretable *at all*, and so cannot be understood as the subject of a proposition. Rather, it appears to be interpretable only in the context of the broader configuration it figures in, which includes the quantifier expression $\forall x$. So we are seemingly left with two options to maintain the (supposedly *a priori*) fundamental organizing principle in this case. Either the grammar of this language is much more complicated than surface appearances suggest, or variables can be reasonably interpreted as subject-designating after all.

What follows is an extended meditation on this particular case study in philosophical grammar. More specifically, it is an exploration of the second of the two options suggested above. I present a view according to which variables are interpreted by a kind of *sui generis* intensional entity. That is to say, variable-meanings are things that may assume different values in different states. This view is directly modeled on the treatment of variables in explicitly compositional formulations of the standard Tarskian semantics of first-order languages. But I adopt a generalized stance towards the states against which variable-meanings are evaluated. This generalized approach allows me to easily define two important notions that are quite difficult to characterize otherwise. First, I can give a semantic formulation of the notion of a *restricted* variable. Intuitively speaking, these are just those variables that only range over some proper sub-region of the universe. So they can be easily modeled by

interpreting them with a variable-meaning that may only assume a value in that range in any given state. And second, I can give a semantic formulation of the notion of a system of *structurally dependent* variables. These are situations in which *sequences* of variables may only assume certain acceptable patterns of values. This is once again easily modeled by interpreting the sequence of variables by a sequence of variable-meanings that can only simultaneously assume an acceptable pattern of values in any given state. The result is a view according to which variables are subject-designating in much the same way that other kinds of intensional expressions are, in the sense that they are officially interpreted by *ways of thinking* about a domain of quantification. What sets variables apart from other types of intensional expression is that these ways of thinking are potentially both plural and structured. In other words, variables express concepts that an agent can use to think about structurally dependent systems of objects in the world.

The first two chapters directly concern the semantic role of variables in first-order languages. Chapter 1 outlines what appear to be two quite different attitudes one might have towards this issue, both of which are largely due to the foundational contributions of Frege and Tarski. Even though their work is essential to our contemporary understanding of logic and semantics, these authors had wildly diverging ways of understanding the basic semantic structure of quantification. According to Frege, the proper semantic treatment of a particular quantifier will appeal to a single object – a higher-order property of some kind. In this context, variables do not function as independently interpretable expressions. Rather, they conventionally signal certain lateral structural relations, in much the same way that other punctuation items conventionally signal certain hierarchical structural relations. Tarski, on the other hand, attempted to accomplish the same truth-conditional effect by making use of at least *two* objects – a sentential operator and a variable that it controls. In this approach, variables are assigned a meaning like any other term. So they serve to make an independent semantic contribution to the compositional computation of the truth-conditions of formulas in which they occur, instead. Each approach appears to have its own benefits and drawbacks,

though ultimately, neither seems to be fully satisfactory.

Chapter 2 resumes this discussion by trying to get a better sense of why the conflict between the two views arises in the first place. Since the difference between the two perspectives is ultimately due to a difference in their underlying grammars, I try to clarify some of the guiding intuitions behind each perspective by more explicitly attending to some of the theoretical roles that the notion of grammatical structure has traditionally been expected to serve. As it turns out, the Tarskian thinks of the structure of a sentence as something like the relevant level of syntactic description for compositional interpretation, whereas the Fregean thinks of it as a level of syntactic description required for the transparent display of logical relations. In general, there is no reason to expect a single notion of grammatical structure will be suitable for adequately serving both of these roles, and I argue that we have good reason to think they come apart in this case. And if a language does not admit a single grammar suitable for both explanatory purposes, it will be necessary to think of the expressions in that language as having multiple structural analyses of quite different kinds. The Fregean and the Tarskian grammar are therefore not really in conflict. Rather, they each fill a theoretical lacuna left by the other. The upshot of this discussion is that the Tarskian perspective is more directly concerned with how humans actually understand quantified first-order statements. And this, in turn, appears to be *instance based* in the sense that quantified sentences are interpreted as statements about the presence or absence of suitable verifying witnesses. So quantified claims turn out to be about the individuals who make up the domains from which witnesses may be drawn, and agents will need to be able to track complex structural dependencies between different objects in those domains. I claim that the Tarskian interpretations of variables model the concepts agents use for that purpose, and so variables may safely be regarded as just another kind of intensional expression.

The remaining chapters concern how this way of analyzing variables can also shed light on the interpretation of pronouns, which are their closest natural language analogs. Chapter 3 reviews some basic facts about the interpretation of anaphoric pronouns in discourse. I claim

that discourses need to be organized in a way that ensures their participants are referentially coordinated, in the rough sense that they are guaranteed to be talking about the same subjects of conversation. I then claim that anaphora's distinctive interpretive role is to help secure this property. One popular strategy for explaining how this is possible begins by noting that certain expression occurrences appear to be governed by singular use-conditions, in the sense that the producers of those occurrences are required to have particular objects in mind to serve as their intended referents. Those intentions can then be used to fix the contents of anaphoric pronoun occurrences. While this view is an extremely simple and attractive starting point, I argue that it will need to be generalized to handle a wider range of cases. In particular, there are many examples of pronoun occurrences that appear to exhibit all of the key interpretive features of anaphora, but where singular use-conditions do not plausibly govern the apparent antecedents of those occurrences.

Chapter 4 discusses a cluster of more general use-conditions that apply to all noun phrase occurrences. In general, the evidential support of an assertion will typically concern some specific individual or collection of individuals. And that evidential support is always conversationally relevant, as speakers are generally expected to be in positions to justify themselves if called upon to do so. Therefore, the producer of a warranted assertion can be expected to have something in mind that can serve as a potential witness to the truth of their statement. I argue that in cases of assertions involving polyadic quantification, these witnesses are best modeled as Tarskian variable-meanings that are both restricted and structurally dependent. We can then get a much more wide-reaching theory of anaphoric relations by taking the contents of anaphoric pronoun occurrences to be fixed by these witness-indicating intentions. The form of referential coordination secured by this view of anaphora is notably stronger than is normally appreciated – conversational participants are seemingly coordinated on potentially complex systems of objects, which may include both individual collections of objects and some non-trivial relations among them.

Since pronouns are context-sensitive expressions *par excellence*, Chapter 5 ties up some

loose ends by showing how this account can be integrated into the standard Kaplanian framework for thinking about context-sensitivity more generally. The defining feature of this framework is a distinction between a context-invariant rule-like notion of meaning and the contextually saturated results of applying those rules on a given occasion. And the ability to make this distinction at the lexical level requires a commitment to a strong form of compositionality. I claim that the account of anaphoric pronouns sketched in the previous chapters can actually serve as the basis for a uniform treatment of all pronoun occurrences in this framework. That means it should be possible to state a single semantic rule that determines the content of any pronoun occurrence in context. While it is more or less straightforward to state this rule, it also implicitly appeals to the idea that context may shift during the evaluation of a given sentence. But context shifts of this kind are normally assumed to be incompatible with the notion of compositionality that the Kaplanian framework seems to require. So I conclude by showing why this standard assumption is incorrect.

I will close this introduction with a couple of brief notes on how this material is presented. The discussion is kept decidedly informal whenever possible to keep the body of the text both legible and accessible. I have therefore included several technical appendices that more explicitly work through some of the distracting details that were consciously omitted. The distinction between use and mention for natural language expressions is systematically observed with the use of teletype font – i.e., cats have four legs, while `cats` has four letters. I am much less vigilant about systematically marking the distinction for expressions in formal languages. But since this is a dissertation written in English, it is safe to assume that any formal expression occurring in it is being mentioned unless it is overwhelming clear from context otherwise.

CHAPTER 1

Two Theories of First-order Quantification

The modern age of logic and semantics began in earnest with a few simple linguistic insights. Frege had noticed that the traditional term logic of his predecessors could be vastly generalized and simplified by reformulating the grammar of certain fundamental logical notions using expressive idioms borrowed from mathematical languages rather than natural ones. And his adaptation and refinement of the informal mathematical practice of using variables to express general truths were undoubtedly some of the more significant reforms he brought about. The apparatus of variables and variable-binding operators that he introduced is still the preferred idiom for representing the structure of complex patterns of quantification. And that fact is unlikely to change any time soon, as it appears to enjoy a kind of practical indispensability. While expressively equivalent variable-free alternative idioms are now well known and well studied, they are undeniably much harder to read and understand and so tend to be deployed only in contexts where concerns about human legibility are at a complete minimum. But despite their pervasiveness, variables presently occupy an awkward position in logical theory. In particular, the philosophical underpinnings of their semantic functioning are still very poorly understood.

It is ordinarily assumed that everything that needed to be said about variables was wholly covered by the pioneering work of Frege and Tarski. But a closer examination of those foundational contributions quickly reveals quite a few apparently irreconcilable tensions. We are left with a seemingly inconsistent and incomplete characterization of a variable's theoretical role. To take some of the first steps towards providing a better characterization

of that role, we first need to cover the basics. Therefore, this chapter aims to bring the conflict between the Tarskian and Fregean views into sharper focus.

In §1, I will briefly review some of the formal details of the standard semantics for first-order languages derived from the work of Tarski. The most distinctive aspect of this approach is that variables are treated as genuine syntactic constituents of expressions. A compositional meaning theory thus has to assign them a semantic value of some kind, and so the Tarskian approach sees variables as individually meaningful.

In §2, I then review some recent criticisms of the Tarskian approach, with special attention paid to Fine (2003, 2007). The unifying theme of these criticisms is that the theoretical role assigned by the Tarskian theory to individual variables is somehow too language-dependent to count as properly semantic. While the Tarskian claims the distinct variables x and y have distinct meanings, the objectors charge that this difference can only be grounded in brute difference in the syntax of x and y. But a properly semantic distinction can only be made on the basis of the more abstract properties that expressions bear to the language-independent reality that they represent.

In §3, I introduce the Fregean analysis of quantification as a way of avoiding some of these pitfalls. This approach is best seen as a kind of eliminativism about variables. In other words, they are not given an independent semantic role to play. Instead, they are seen as a kind of punctuation device, serving to broadcast the non-local syntactic relation that holds between a quantifier and the positions in a complex predicate that it controls. But since Frege did not supply an explicit compositional semantics for his language in the way that we are now accustomed to, I instead focus on the recent formal implementation of Frege's semantic suggestions offered by Wehmeier (2018).

I conclude in §4 with some novel criticisms of the Fregean theory. The unifying theme of my criticisms is that the Fregean theory is not positioned to explain certain essential facts about the use of first-order languages that a semantic theory ought to explain. I first argue that the formal account of the Fregean perspective fails because it is either unlearnable or

relies on an untenably weak characterization of what a speaker must know to understand the use of a constant symbol. I then give a more general argument that any Fregean account (regardless of formal implementation) will be unable to account for the fact that human speakers appear to interpret first-order claims incrementally.

1.1 Basics of Tarskian Quantification

The contemporary understanding of the formal semantics of quantification in first-order languages is directly derived from Tarski (1935). In that work, Tarski showed how it was possible to rigorously define the concept of truth in a broad family of formal languages. To avoid the semantic paradoxes, these definitions must always be understood as being given in a metalanguage whose expressive resources are strictly richer than those of the object language for which truth is being defined. The result was a general recipe for giving scientifically acceptable definitions of relativized truth predicates (i.e., true-in-$\mathcal{L}$). This recipe assumes that the language it is being applied to is structurally quite straightforward. In particular, it needs to be possible to isolate a class of atomic formulas, and any non-atomic formula needs to be uniquely analyzable as the result of applying some syntactic operation to simpler formulas. This allows the relativized truth predicate to be recursively defined by directly specifying when atomic formulas are true and by specifying the truth of complex formulas in terms of the truth of their simpler components.

For our purposes, what matters is how this general recipe was used for the quantifiers. Intuitively speaking, a formula $\exists x Fx$ ought to be true just in case there is some value that x may take such that Fx is true. In other words, the truth of a general formula is seemingly reducible to the truth of its (simpler) possible singular instances. There are several different ways that this basic underlying idea can be formalized. For instance, a substitutional approach would hold that $\exists x Fx$ is true just in case there is a closed term t such that Ft is true. But every object in the intended domain of quantification needs to

be denoted by some closed term in the language for this strategy to yield the correct truth conditions. And as there is no guarantee that this is, in fact, the case, it does not seem to be the best way to proceed in general.[1] Instead, Tarski chose to further relativize the truth-in-$\mathcal{L}$ predicate to an additional parameter that explicitly formalized the idea that a variable could potentially have any object in the domain of quantification as its value. Each variable can be seen as designating a definite object with respect to this parameter. The correct characterization of the semantic behavior of object-language quantification is then successfully captured by meta-language quantification over this new parameter.

More explicitly, we assume that we are given a fixed language $\mathcal{L}$, which for simplicity, we may assume consists of a set $\mathcal{C}$ of constant symbols, and a set $\mathcal{P}$ of predicate symbols equipped with an arity function $\rho : \mathcal{P} \to \mathbb{N}$ that tells us how many arguments each predicate symbol expects. In addition, we will assume that we are given a fixed countably infinite stock of variables $\mathcal{V}$. Atomic formulas are constructed by concatenating a predicate $P \in \mathcal{P}$ with an appropriate number ($\rho(P)$) of variables or constants. More complex formulas can be constructed from simpler ones by means of truth-functional connectives ($\neg, \wedge$) or quantifier phrases ($\exists v, v \in \mathcal{V}$). More precisely, the formulas of predicate logic are precisely those strings generated by the following context-free grammar:

[1]There are quite a few ways around this apparent problem other than the one discussed here. For instance, it is now a reasonably common practice to first expand languages with enough constants to ensure that the substitutional truth conditions are correct before giving an official definition of truth. And as a more conservative variant of this proposal, Lavine (2000) (building on some suggestions in Geach (1980)), shows how this substitutional approach can likewise be salvaged utilizing finite language expansions that get made as they are needed, rather than all at once. While these substitutional approaches are sometimes thought to hold certain conceptual advantages over the standard Tarskian one presented in this section (see, e.g., Marcus (1972)), it turns out to be extremely difficult to reformulate these proposals as compositional meaning assignments while maintaining reasonable assumptions about syntactic form (see Janssen (1997) for a brief discussion of this point). This fact makes them extremely inappropriate for the task at hand, and so they will be mostly ignored in what follows.

Tarskian Grammar of First-order Logic:

$$T \longmapsto c \mid v \ (c \in \mathcal{C},\, v \in \mathcal{V})$$
$$Var \longmapsto v \ (v \in \mathcal{V})$$
$$Pred_n \longmapsto P \ (P \in \mathcal{P} \text{ and } \rho(P) = n)$$
$$S \longmapsto Pred_n \underbrace{T \ldots T}_{n\text{-times}} \mid \neg S \mid (S \wedge S) \mid \exists\, Var\ S$$

Formulas are evaluated with respect to a model $\mathcal{M}$, and an assignment of values in $\mathcal{M}$ to variables.[2] A model $\mathcal{M} = \langle D, I\rangle$ is a pair of a non-empty set D (the domain), and an interpretation function I that maps each constant symbol to some element in D and each n-ary predicate symbol to some n-ary relation on D. For any particular model $\mathcal{M} = \langle D, I\rangle$, an assignment of values in $\mathcal{M}$ to variables is any function $g : \mathcal{V} \to D$. Two variable assignments g, g' are said to be v-variants (written $g \sim_v g'$) just in case $g(v') = g'(v')$ for all variables $v \neq v'$. The denotation of a term t with respect to a variable assignment g and model $\mathcal{M}$ ($den_g^{\mathcal{M}} t$)) is just $I(t)$ when $t \in \mathcal{C}$, and $g(t)$ otherwise. The recursive definition of truth with respect to a model and assignment then runs as follows:

Tarskian Definition of Truth

$$\begin{array}{lll}
\mathcal{M}, g \vDash Pt_1 \ldots t_n & \text{iff} & \langle den_g^{\mathcal{M}}(t_1), \ldots, den_g^{\mathcal{M}}(t_n)\rangle \in I(P) \\
\mathcal{M}, g \vDash \neg\phi & \text{iff} & \mathcal{M}, g \nvDash \phi \\
\mathcal{M}, g \vDash (\phi \wedge \psi) & \text{iff} & \mathcal{M}, g \vDash \phi \text{ and } \mathcal{M}, g \vDash \psi \\
\mathcal{M}, g \vDash \exists v\phi & \text{iff} & \text{for some } g' \sim_v g,\ \mathcal{M}, g' \vDash \phi
\end{array}$$

Finally, truth in a model (simpliciter) is defined in terms of truth in a model under an assignment – ϕ is true in $\mathcal{M}$ ($\mathcal{M} \vDash \phi$) just in case for all variable assignments g, we have $\mathcal{M}, g \vDash \phi$.

[2]The addition of a model is not an actual departure from Tarski (1935). The languages treated in that initial work were assumed to have a fixed interpretation, but it has proved to be much more convenient to see that interpretation as potentially varying. Tarski's definition of truth-in-$\mathcal{L}$ becomes a definition of truth-with-respect-to-$\mathcal{M}$ once that change of perspective is adopted.

While the above definition of truth (or some variant of it) is what one in fact works with in most applications of first-order logic, it does not take the form that a formal semantics is now generally assumed to take in mainstream linguistic semantics and philosophy of language. Semantic theories are expected to conform to the *principle of compositionality*. In crude terms, this principle states that the meaning of a complex expression is functionally determined by the meanings of its parts and their syntactic arrangement. In order to ensure that a given semantic theory in fact conforms to this principle, it is standardly required that it take the form of an explicit *compositional meaning assignment*. These directly specify the semantic values of certain basic expressions and then give a handful of rules for computing the semantic values of larger expressions on the basis of the semantic values of their parts (typically one for each mode of syntactic combination). And a recursive definition of truth does not conform to the principle as it is standardly stated because (among other things) it simply does not proceed by associating semantic values to expressions.

Of course, this is not a major conceptual problem by itself. Most of the familiar arguments for compositionality stem from worries about the cognitive limitations of language users. For instance, it is often claimed that languages that are not compositional would be difficult or impossible for an agent with bounded cognitive resources to learn. Likewise, the apparent fact that resource-bounded agents can produce and interpret an unbounded number of novel utterances is taken to be best explained by assuming the language in question is compositional.[3] But if the notion of meaning can be understood in terms of the notion of truth, then a recursive definition of truth in an interpreted language (i.e., truth with respect to a fixed model $\mathcal{M}$) seems no worse than a compositional meaning assignment for explaining these cognitive facts. In fact, the only real difference between the two kinds of approaches is whether or not individual semantic values (meanings) are systematically associated with expressions. This makes recursive definitions of truth the preferred semantic framework to

[3]See Pagin and Westerståhl (2010) for a recent overview of arguments for compositionality.

work in for those who are somehow skeptical of reified meanings.[4] However, since my present concern is the proper semantic analysis of certain sub-sentential expressions (quantifiers and variables), it will be convenient to be able to talk as though these expressions were assigned some definite meanings by the theory. And since I am not particularly skeptical of reified meanings, the easiest way to facilitate this way of talking will be to just explicitly reformulate the basics of the Tarskian theory as a compositional meaning assignment instead.

Thankfully, such a reformulation is not particularly hard to come by.[5] To begin, note that the clauses for atomic expressions and the propositional connectives in the standard definition of truth are still compositional in an intuitive sense. They tell us how to compute the value of an expression with respect to the two relevant parameters of evaluation on the basis of the values of its immediate sub-expressions at those very same parameters. But the clause for the universal quantifier requires that we vary the assignment parameter, and that means that the value of a complex expression is not obviously determined by the value of its immediate components. The essence of this problem is that the obvious candidate for the meaning of a formula under an assignment (its truth-value) does not encode enough information in order to determine what the appropriate semantic action of the universal quantifier should be. Hence, we need a richer notion of meaning that does carry enough information to get the job done. And since the problem was caused by the quantifier clause's need to vary the assignment parameter, the obvious solution is to internalize that parameter and take a formula's meaning to be the set of its satisfying assignments.[6]

---

[4]See e.g. Quine (1960b, 1970) and Davidson (1965).

[5]While extremely rare, it is not even completely unprecedented for a textbook to give the semantics of predicate logic an explicitly compositional form – see, e.g., Kreisel and Krivine (1967) or Monk (1976). This approach is based on Tarski's work trying to find a workable algebraic semantics of predicate logic. Just as Boolean algebras are convenient structures to use in order to model classical propositional logic, it turns out that *cylindric algebras* are feasible algebraic analogs of classical predicate logic when the Tarskian grammar is held fixed. Restricting our attention to formulas, the compositional semantics we are about to give is the paradigm concrete realization of a structure of this kind. All of this is just to say that while this semantics is not normally what one associates with Tarski's name, it is equally deserving of the title in the discussion that follows. See Henkin et al. (1971) for a standard reference on cylindric algebras.

[6]Formally, this is exactly analogous to what happens when we intensionalize the semantics of propositional

The compositional reformulation of the Tarskian theory along these lines runs as follows. Fixing a background model $\mathcal{M} = \langle D, I \rangle$, we first need to give a lexicon that assigns each basic expression e its semantic value in that model $[\![e]\!]$:[7]

Tarskian Compositional Lexicon

Expression (e)	Semantic Value ($[\![e]\!]$)
c $[c \in \mathcal{C}]$	$\lambda g.I(c)$
v $[v \in \mathcal{V}]$	$\lambda g.g(v)$
P $[P \in Pr$ and $\rho(P) = n]$	$\lambda x_1 \ldots \lambda x_n.\{g \mid \langle x_1(g), \ldots, x_n(g)\rangle \in I(P)\}$
$\wedge$	$\lambda\phi.\lambda\psi.\phi \cap \psi$
$\neg$	$\lambda\phi.D^{\mathcal{V}} \setminus \phi$
$\exists$	$\lambda x.\lambda\phi.\{h \mid g \in \phi \;\&\; h \sim_x g\}$

We then need to specify a family of operations that can be used to compute the semantic values of complex expressions on the basis of their simpler parts. This is especially straightforward given how we have chosen to design our lexicon. In particular, syntactic combination can always be interpreted as a form of function application. More explicitly:

logic in order to also be able to account for the meanings of modal operators. Since the appropriate notion of sentential meaning in propositional logic (again, a truth-value) is not rich enough to handle non-truth-functional connectives, we need to promote a formerly backgrounded evaluation parameter (the valuation-function, now under the guise of a possible world) and take the meanings of sentences to be sets of such parameters.

[7]There is actually a slight abuse of notation in the clause for the existential quantifier. Normally, $\sim_x$ is thought of as being parameterized by a piece of syntax (the actual variable x). But here, you are asked to think of it as being parameterized by the corresponding meaning of the variable. Note also that now (technically) the existential quantifier is well-defined for constant inputs, in which case it can be stipulated to return the identity function on $\wp(D^{\mathcal{V}})$. We do not need to worry about this case, however, since this kind of combination will never be syntactically well-formed.

Tarskian Composition Rules

$$\begin{aligned}
[\![Pt_1 \ldots t_n]\!] &= [\![P]\!]([\![t_1]\!], \ldots, [\![t_n]\!]) \\
[\![\neg\phi]\!] &= [\![\neg]\!]([\![\phi]\!]) \\
[\![(\phi \wedge \psi)]\!] &= [\![\wedge]\!]([\![\phi]\!], [\![\psi]\!]) \\
[\![\exists x\phi]\!] &= [\![\exists]\!]([\![x]\!], [\![\phi]\!])
\end{aligned}$$

Taken together, the above lexicon and composition rules ensure that the semantic value of a sentence is always its set of satisfying sequences. So since we used to say that a sentence ϕ is true in $\mathcal{M}$ just in case $\mathcal{M}, g \vDash \phi$ for all assignments g, we should now instead say that ϕ is true in $\mathcal{M}$ just in case $[\![\phi]\!] = D^{\mathcal{V}}$.

With these formal details in place, it is now possible to isolate the most distinctive feature of the Tarskian approach to the semantics of quantification. First, note that the language we have applied this approach to recognizes open formulas as grammatically well-formed expressions. Because variables are treated as genuine terms, and nothing in the grammar itself cares about the distinction between free and bound occurrences of variables, open formulas are structurally indistinguishable from their closed counterparts. By itself, this observation is not particularly interesting. But since the semantics for this language tightly hews to its grammar, this boring point about structure becomes a much more important point about meaning. Specifically, open formulas are treated as being semantically as well as grammatically legitimate. But is there any intuitive sense that can be made out of the Tarskian meaning of an open formula?

In order to understand an open formula's characteristic semantic contribution, we need to consider the role that it plays in the analysis of a general statement. Remember that the Tarskian approach to quantification is an attempt to formalize the underlying idea that the existential statement $\exists xFx$ should be true just in case there is a possible value that x may assume such that Fx is true. So, according to this approach, the truth of a general statement boils down to a claim about possible designation – i.e., it says that it is possible

for x to designate an object such that Fx is true. When stated in these terms, it seems like we can straightforwardly understand the semantics of variables and open formulas on a kind of analogy with well-established ways of talking about intensional phenomena more generally. Just as the definite description `the greatest living logician` is a non-rigid designator whose denotation may vary with respect to possibilities of various kinds (in this case, world/time pairs), the variable x in a Tarskian framework is also a non-rigid designator of a certain kind. Formally speaking, it is just an expression whose denotation is allowed to vary along a certain dimension. In other words, a variable's denotation varies depending on the state that it is in. And the fact that a variable's denotation depends on its state means that any larger expression that contains that variable may also depend on its state in a similar fashion. The resulting picture does not appear to meaningfully depart from otherwise normal practices in semantic theory. Just as propositions are typically identified with sets of possibilities, so too are the meanings of open formulas identified with sets of possibilities in the Tarskian framework (where these possibilities characterize the different states that each of the variables can be in).

1.2 Problems with Tarskian Quantification

Unfortunately, this way of understanding the Tarskian approach has struck many as deeply problematic. A typical complaint leveled against this strategy is that the resulting space of semantic values is too language-dependent to count as genuinely semantic. Instead of being made from what we would take to be normal model-theoretic objects, meanings are constructed from assignments, and assignments are in turn constructed from a set of variables $\mathcal{V}$. But the actual choice of $\mathcal{V}$ is an entirely conventional aspect of our syntax. So if part of the point of giving a compositional meaning assignment is to specify how formal expressions are related to a language-independent subject matter, then we have a serious problem. Meanings

and the expressions that are used to express them are far too closely related.[8]

This worry can be made a bit sharper by considering one of its more disquieting consequences. Since the set of variables that a language uses is completely arbitrary, the choice of that set should not change the range of things that you can say. In particular, it should be possible to easily translate between two languages that use disjoint sets of variables. Now, one of the most cited theoretical roles that a notion of meaning is supposed to play is to serve as a kind of invariant under a good translation procedure of this kind. That is to say, a minimum requirement for a given translation to count as a good one is that the translated expression has the same meaning as its translation. But if meanings partially bottom out in facts that depend on the set we chose at the outset, then it will, in general, be impossible to translate between languages that use disjoint sets of variables in a way that meets this general condition.

Of course, these remarks as they stand are not sufficient to establish a genuine problem. After all, it is a relatively trivial matter to remove the explicit presence of syntax in our space of semantic values by trading variable assignments for ω-sequences of objects from the underlying domain. The switch only involves an extremely minor requirement that we also have some fixed enumeration of the set of variables $\mathcal{V}$. We can then take $[\![v_i]\!]$ to be the i-th projection function and say that $s \sim_{v_i} s'$ iff s differs from s' at most with respect to its i-th component.[9] On this way of doing things, there would not be any obviously non-semantic entities floating around in our models. And since the set of variables we choose and how we choose to enumerate its elements do not make any principled difference to what we can express in the language, the possibility of an adequate translation between languages that

[8]This worry is now extremely prevalent in the literature. Wehmeier (2018) attributes it primarily to Zimmermann and Sternefeld (2013) and Klein and Sternefeld (2017), but concerns of this kind are also part of the motivation behind Jabcobson's variable-free program (see e.g. Jacobson (1999, 2003)). It also seems to be implicit in more formally-oriented approaches to discourse representation theory. See, e.g., Vermeulen (1995) and Muskens (1996), which are both extremely careful to distinguish between variable names and the model-theoretic objects they pick out.

[9]This is in fact how many presentations of the basic semantics first-order languages proceeds in the first place, including the original presentation in Tarski (1935).

make use of different sets of variables now looks perfectly achievable in a very straightforward way.

Unfortunately, Fine (2003, 2007) can be seen as giving a subtler and more sophisticated version of this worry that is not as easily dismissed. Instead of appealing to the apparent semantic equivalence of expressions in different languages, Fine appeals to the apparent semantic equivalence of expressions in the same language. This worry is presented in the form of a puzzle that he calls the *antinomy of the variable*, which he attributes to Russell (1903). To a first approximation, it runs as follows: we are generally inclined to say apparently contradictory things about the semantic roles that the variables x and y play. In certain situations, they appear to be the same. Whatever you want to say about the role in speech that the assertion of a statement with a free variable is supposed to serve, it seems clear enough that $x > 0$ and $y > 0$ are about as theoretically indistinguishable as you can get – they are mere orthographic variants of one another. So x and y appear to make exactly the same semantic contribution in this environment. But in other situations, this role appears to be quite different. Proper understanding of the open formula $x > y$, whatever that is supposed to consist in, seems to at least minimally involve recognizing that x and y are playing different semantic roles despite ranging over the same class of objects. Why else would $x > x$ differ in meaning from $x > y$?

We can make Fine's point a little more explicit as follows. Without getting too bogged down in the details, we define a syntactic environment $\phi[\xi_1, \ldots, \xi_n]$ to be a formula with n distinct 'gaps'. Given a sequence of expressions $\langle e_1, \ldots, e_n \rangle$ and a syntactic environment $\phi[\xi_1, \ldots, \xi_n]$, $\phi[e_1, \ldots, e_n]$ is the result of 'plugging' each gap ξ_i with e_i. Call two sequences of expressions $\vec{e}$ and $\vec{e'}$ *environmentally distinguishable* just in case there is a syntactic environment $\phi[\vec{\xi}]$ such that (i) $\phi[\vec{\xi}]$ does not contain an occurrence of an expression that also occurs in either $\vec{e}$ or $\vec{e'}$ and (ii) $\phi[\vec{e}]$ says something different than $\phi[\vec{e'}]$. Our conflicting intuitions can then be summarized by the following two observations:

(**Sameness**) The sequences $\langle x \rangle$ and $\langle y \rangle$ are not environmentally distinguishable.

(**Difference**) The sequences $\langle x, x\rangle$ and $\langle x, y\rangle$ are environmentally distinguishable.

Note that this refinement of the puzzle requires that we understand the notion of semantic role in a certain specific way. In particular, the semantic role of an expression must be characterized in terms of the role that it plays in communication. In other words, a difference in semantic role needs to correspond to a difference in the contribution that it makes to what is communicated in literal discourse.[10] If we think of syntactic environments as encompassing what would otherwise be a complete communicative episode if not for the gaps, then it seems hard to deny either (**Sameness**) or (**Difference**). (**Difference**) is easily established by just noting that the environment $\xi_1 > \xi_2$ suffices to distinguish $\langle x, x\rangle$, and $\langle x, y\rangle$ (one way of filling in the gaps results in something necessarily false, while the other is perfectly consistent). And (**Sameness**) follows from our apparent indifference to the choice between equivalent alphabetic variants of a given expression.

As stated in these terms, there is no actual contradiction between these two observations. But there is an undeniable tension between them that is made much more palpable once we start to make additional assumptions about the relationship that a given semantic theory needs to have to the notion of semantic role at work in the antinomy. More specifically, Fine seems to think that if two expressions e_1 and e_2 have the same semantic role, then an adequate semantic theory will need to assign them the same semantic value (i.e., it will have to yield the judgment that $[\![e_1]\!] = [\![e_2]\!]$). And since environmental indistinguishability is our main test for the sameness of semantic role, it follows that a semantic of variables has to yield the judgment that $[\![x]\!] = [\![y]\!]$. But this means that compositionality will fail in its most familiar form since an adequate semantic theory will also clearly need to yield the judgment that $[\![x > x]\!] \neq [\![x > y]\!]$. So the dilemma that the antinomy leaves us with is the following: a semantics of variables must either be non-compositional, or it must recognize

[10] While Fine says surprisingly little about this topic, it is clear from context that this is how he understands the notion.

more distinctions between semantic values than those that can be motivated on the basis of semantic role alone.[11]

As the Tarskian theory presented in the previous section is compositional, it clearly goes in for the second horn of this dilemma. So, according to Fine, the notion of semantic value at work makes certain *purely* formal distinctions that are not obviously borne out in the language's manifest patterns of use. This problematic aspect of the Tarskian proposal can be more directly illustrated in the following way. Suppose that the semantic value of an open formula in the Tarskian theory was something that could play a reasonable role in communication. In other words, suppose that we could identify them with assertable contents of some kind. Then just note that generally speaking, an assertable content will encode the information about which specific variables need to be used in order to assert it.[12] Thus, meanings still depend on language in a certain sense – assertable contents are always interpretable as being 'about' the values that certain specific items of syntax (i.e., the variables that are required to assert it) are capable of jointly assuming. Thus, *no* compositional semantics for a first-order language that countenances free variables is going to be sufficiently abstract. It will always contain some kind of mirror of syntax that is not directly reflected in the non-linguistic reality that the language is used to describe.

[11]Fine goes in for the first horn of the dilemma. According to his *semantic relationism*, semantic values can only be directly assigned to sequences of expressions, so that we may have $[\![x]\!] = [\![y]\!]$ while also having $[\![\langle x, x\rangle]\!] \neq [\![\langle x, y\rangle]\!]$. And as the semantic value of a sequence of expressions is not in general derivable from the semantic values of its components, this amounts to a denial of compositionality in some important sense (but see Pickel and Rabern (2017) for a discussion of whether or not Fine's theory is reasonably regarded as non-compositional, or as merely positing a more sophisticated background syntax.)

[12]Pick any formula $\varphi(\vec{x})$ and let σ be a sequence that satisfies it. Then every sequence σ' that differs from σ only in positions not associated to the variables $\vec{x}$ *also* satisfies $\varphi(\vec{x})$. Hence $\pi_i([\![\varphi(\vec{x})]\!])$ will fail to be the entire domain only for finitely many choices of positions i. The positions that fail to be the entire domain thus tells us precisely which variables we *must* use in order to express $[\![\varphi(\vec{x})]\!]$.

1.3 Basics of Fregean Quantification

A common response to the cluster of issues discussed in the previous section is to simply reject the core tenet of the Tarskian theory and deny that open formulas are objects of serious semantic inquiry. While there are a number of different ways of developing this basic underlying idea, most generally involve abandoning the variable-binding idiom for the expression of generality altogether.[13] But a notable exception to that general rule can, in fact, be found in one of the earliest discussions of that idiom. In particular, the remarks in the earlier sections of Frege (1893) provide what seems to be an outline for how to think about the semantics of this language without the need to appeal to variables at all.

Frege's discussion has a very elementary type theory working in the background. For our purposes, it suffices to note that he makes a distinction between *objects* on the one hand and *functions* on the other. Objects are *saturated*, while functions are *unsaturated*. Functions may then be further subclassified into the kinds of things that they need to be supplied with in order to become saturated. For instance, a *first-level* function needs an object, while a *second-level* function needs a first-level function.[14] The need for something unsaturated to be supplied with some additional component in order to become saturated is then taken to be the main driving force behind semantic composition. And since the form of the principle of compositionality that Frege appeared to adhere to required that syntactic and semantic composition closely mirror each other, these basic type-theoretic assumptions about the semantics of his logical language informed some of his highly idiosyncratic views about its grammar. According to Frege, every expression in his formal language is a *name* for something, including sentences (which are names for truth-values). Given a lexicon of

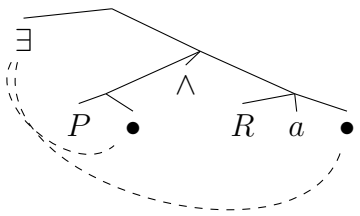
[13] Quine's predicate functor calculus (Quine (1960a)) is one attempt to eliminate variables from first-order logic. And more generally, the well-known equivalence that holds between combinatory logic and the λ-calculus can be exploited to secure the elimination of variables in any higher-order setting. See e.g. Hindley and Seldin (2008) for more details.

[14] You can obviously keep going in this fashion, and the full language that Frege describes includes primitives for *third-level* functions as well in order to treat higher-order quantification, but since we are restricting our attention to his views on first-order quantification here, we can safely stop at the second level.

primitive names, you may form a new name in one of two distinct ways. First, you may generate a new name for an object by first applying a name of a function to the name of something of the appropriate type.[15] And second, you may generate a new name for a first-level function by taking a name for an object and replacing one or more occurrences of a name for an object with a gap. For instance, the (gappy) expression $(P\xi \wedge Ra\xi)$ is a name for a first-level function that is derivable from a name for a truth-value like $(Pb \wedge Rab)$.

This way of thinking about the syntax and semantics of first-order languages seems to circumvent the need for an independent semantics of variables altogether. For instance, Frege would analyze $\exists x(Px \wedge Rax)$ as the result of applying the (primitive) name of a second-level function $\exists$ to the (derived) name of a first-level function $(P\xi \wedge Ra\xi)$. For the Fregean, variables syntactically function as pieces of uninterpreted punctuation. They serve the grammatical role of indicating lateral quantifier-control structure, much like parentheses serve the grammatical role of indicating hierarchical constituent structure. More explicitly:

(1) a. $\exists x(Px \wedge Rax)$

b. 

This line of thought holds that the formula in (1a) has something like the grammatical structure that is displayed in (1b). If this is right, the occurrences of the variable x make no independent interpretive contributions – instead, they merely serve to telegraph the structural relation that holds between a quantifier and the positions in a complex predicate that

[15] In other words, you can generate a new name for an object by applying the name of a first-level function to the name of an object or by applying the name of a second-level function to the name of a first-level function.

it governs.[16]

Of course, the success of a project of this kind obviously depends on our ability to provide an adequate interpretation of Frege's second syntactic rule for generating new names. In particular, we need to explain how the meanings of (gappy) expressions like $(P\xi \wedge Ra\xi)$ are derived from the meanings of their closed instances like $(Pb \wedge Rab)$. Unfortunately, Frege does not explicitly state how this should be done, and it is not clear that it can be done in a way that is compatible with some of his other semantic commitments.[17] But it will still be a useful exercise to walk through some of the details of a more fully worked-out proposal along these lines to put some of Frege's ideas into sharper relief, even if that proposal must ultimately be rejected as inadequate for some reason or other. The rest of this section is therefore dedicated to a quick exposition of some of the main details of Wehmeier (2018), which is the most formally explicit presentation of the Fregean perspective that I am presently aware of.

To begin, note that since the Fregean wants to deny the meaningfulness of open sentences, their grammar needs to officially generate just the closed sentences of first-order logic. But an easy application of the pumping lemma for context-free languages shows that this language is not context-free.[18] So we cannot describe this language using the familiar kind of phrase-structure grammar used to describe the Tarskian language. Another important point of divergence from the Tarskian theory is that we now must require that the stock of constant symbols $\mathcal{C}$ be infinite. This is syntactically necessary since the derivation of a sentence with

[16]For more on this general line of thought, see the remarks in Quine (1951), §12, and Kaplan (1986) regarding the equivalence of the standard variable-binding apparatus to an alternative variable-free notation that instead makes use of coordinating wires.

[17]For instance, it is clearly not possible to systematically derive the function that the gappy expression $(P\xi \wedge Ra\xi)$ names from the mere truth-value that one of its closed instances like $(Pb \wedge Rab)$ (alone) names. And this looks like it might conflict with Frege (1892)'s commitment to the principle of compositionality at the level of reference in addition to sense.

[18]See Partee and Marsh (1987). Note that this argument depends on a certain (extremely natural) mechanism for generating the set of variables, which is a necessary but distracting technical nicety that my discussion has ignored.

n distinct variables will require the abstraction of n distinct constants, and in principle, n can be as large as we would like. And, as we will see, it is also semantically necessary since we will essentially use the portion of a model that interprets the infinite sequence of constants to emulate the behavior of a variable assignment in giving the semantics for condition abstraction.[19]

With this in mind, we can give a simultaneous recursive definition of Fregean sentences and conditions as follows:

Fregean Grammar of First-order Logic

1. Whenever $\rho(P) = n$ and $c_1 \ldots c_n$ are constants, $Pc_1 \ldots c_n$ is a sentence
2. Whenever ϕ and ψ are sentences, so are $\neg\phi$ and $(\phi \wedge \psi)$
3. Whenever ϕ is a sentence with at least one occurrence of the constant c, the result $\phi_c[\xi]$ of replacing every occurrence of c with a blank space ξ is a condition.
4. Whenever δ is a condition, and x is a variable that does not occur in δ, the result $\exists x\delta[x]$ of simultaneous prefixing δ with $\exists x$ and filling all blank spaces with the variable x is a sentence.

Models for Fregean sentences are just as they were in the Tarskian case. Given a constant c and two models $\mathcal{M}$ and $\mathcal{N}$, we say that $\mathcal{M}$ is a c-variant of $\mathcal{N}$ just in case $\mathcal{M}$ and $\mathcal{N}$ are exactly alike except for perhaps how they interpret c. In other words, they share the same domain and interpret all predicates and names other than c in exactly the same manner. Given a constant c, a model $\mathcal{M}$, and an element from its underlying domain u, we let $\mathcal{M}[c := u]$ be the unique c-variant of $\mathcal{M}$ such that $\mathcal{I}(c) = u$. With these terminological

[19]More accurately, the syntactic need for this assumption seems to be more-or-less subservient to the semantic need. Syntactically, it can be easily avoided if we allow for *partial* abstraction of constants (i.e., removing some but not all occurrences of a constant in the formation of a condition). But a natural semantics for partial abstraction is not obviously possible without sacrificing some other quite natural Fregean theoretical desiderata.

preliminaries in place, we can compute the truth-value of a sentence in a model directly as follows:

Fregean Definition of Truth

$$\begin{aligned}
[\![Pc_1 \dots c_n]\!]^{\mathcal{M}} &= 1 \text{ iff } \langle \mathcal{I}(c_1), \dots, \mathcal{I}(c_n) \rangle \in I(P) \\
[\![\neg\phi]\!]^{\mathcal{M}} &= 1 - [\![\phi]\!]^{\mathcal{M}} \\
[\![\phi \wedge \psi]\!]^{\mathcal{M}} &= min([\![\phi]\!]^{\mathcal{M}}, [\![\psi]\!]^{\mathcal{M}}) \\
[\![\phi_c[\xi]]\!]^{\mathcal{M}} &= \{u \in D \mid [\![\phi]\!]^{\mathcal{M}[c:=u]} = 1\} \\
[\![\exists x \delta[x]]\!]^{\mathcal{M}} &= 1 \text{ iff } [\![\delta]\!]^{\mathcal{M}} \neq \emptyset
\end{aligned}$$

Notice, however, that this computation (much like the initial Tarskian one) does not constitute a compositional meaning assignment. The issue, unsurprisingly, has to do with the clause for condition abstraction. Once again, $[\![\phi]\!]^{\mathcal{M}}$ (a truth value) does not carry enough information to uniquely determine what the semantic result of abstracting out a particular constant symbol should be. Luckily, we can once again avoid this problem by applying the same technical trick we used for the Tarskian semantics and lift the meaning of sentences from truth-values to sets of evaluation parameters. This time, however, the only relevant parameter is the background model, so our compositional version of this semantics will end up assigning sets of satisfying models as semantic values to sentences, which can usefully be thought of as their truth conditions. Officially:

Fregean Compositional Meaning Assignment

$$\begin{aligned}
[\![Pc_1 \dots c_n]\!] &= \{\mathcal{M} \mid [\![Pc_1 \dots c_n]\!]^{\mathcal{M}} = 1\} \\
[\![\neg\phi]\!] &= \{\mathcal{M} \mid \mathcal{M} \notin [\![\phi]\!]\} \\
[\![\phi \wedge \psi]\!] &= [\![\phi]\!] \cap [\![\psi]\!] \\
[\![\phi_c[\xi]]\!] &= \lambda\mathcal{M}.\{u \in dom(\mathcal{M}) \mid \mathcal{M}[c := u] \in [\![\phi]\!]\} \\
[\![\exists x \delta[x]]\!] &= \{\mathcal{M} \mid [\![\delta]\!](\mathcal{M}) \neq \emptyset\}
\end{aligned}$$

While there are still many respects in which this proposal could be cleaned up and improved,

for our purposes, it will serve as an adequate formalization of the general Fregean idea.[20]

At this point, I should point out that this theory does not obviously evade all complaints of syntactic pollution. Sentences in this Fregean framework now code information about which constants must be used in order to express them.[21] But this may still be seen as an improvement over the situation that the Tarskian was in. After all, a thoroughgoing Fregean outlook will presumably allow (and maybe even require) that distinct primitive constants be assigned distinct senses. And we should expect these distinctions to project upwards to the level of sentential senses as well. Hence, this mirror of syntax will probably be more palatable (and perhaps even principled) to those with sympathies in line with the other aspects of Frege's semantic machinery, so long as we take this semantics to be formulated at the higher of his two semantic levels.

1.4 Problems with Fregean Quantification

While this Fregean proposal may do better than the Tarskian one with respect to Fine's criticisms, it has certain problems of its own that remove it from contention as a serious alternative. As I mentioned earlier, most of the traditional arguments for compositionality concern the cognitive limitations of language users. While it is not entirely obvious what specific conclusions we may draw from these arguments, they at least show that (all else being equal) one of the major burdens of a semantic theory is to explain how agents with limited computational resources can grasp an infinite number of expression/meaning pairs. And the Fregean of Wehmeier's mold is not capable of shouldering that burden.

---

[20]Due to several technical and conceptual issues that do not concern us here, this is not the final proposal of Wehmeier (2018). But my general complaints against this version of the theory apply equally well to the more complicated version he eventually endorses.

[21]To see this, let $[\![\phi_c[\xi]]\!]$ be a non-trivial condition – that is to say, there is some $\mathcal{M} = \langle D, I\rangle$ such that $[\![\phi_c[\xi]]\!](\mathcal{M}) \neq D$ and $[\![\phi_c[\xi]]\!](\mathcal{M}) \neq \emptyset$. Without loss of generality, suppose $\mathcal{M} \in [\![\phi(c)]\!]$ and let d be a constant not occurring in $\phi(c)$. By assumption, there is some $a \notin [\![\phi_c[\xi]]\!](\mathcal{M})$, and hence $\mathcal{M}[d := a] \notin [\![\phi(d)]\!]$. But clearly $\mathcal{M}[d := a] \in [\![\phi(c)]\!]$, and hence $[\![\phi(c)]\!] \neq [\![\phi(d)]\!]$. Thus, for any non-trivial claims, a difference in the choice of constants corresponds to a difference in meaning.

The key technical trick that Wehmeier exploits is an obvious informational equivalence that holds between model/assignment pairs on the one hand and a certain special class of models on the other. To see what I mean by this, consider the limiting case where the language in question is constant-free. A model of this language would just specify a domain and how to interpret whatever predicates the language happens to have. In this context, the work done by an additional variable-assignment parameter is indispensable – the information that it carries cannot be obviously replicated by any part of the underlying models. But notice how one can trivially promote a model/assignment pair of this kind into a straightforward model of the same language extended by countably many constants. So we could easily get away with specifying models of the former language with models of the latter. But this should not surprise anyone since it is really just a notational variant of the familiar model and assignment apparatus. The interesting observation comes when we consider the other end of the spectrum, when the language has infinitely many constants. In a normal Tarskian setup, any particular computation of meaning is going to only make use of a finite segment of the data furnished by the assignment parameter. Likewise, only a finite segment of the data provided by the model will be used to interpret constant symbols. Given that both sources of data are infinite streams, so to speak, the possibility of folding them together into a single source immediately suggests itself. If we can design the syntax and semantics of the language so as to conspire to always keep the portion of the stream responsible for interpreting occurrences of constants and the portion of the stream responsible for interpreting quantificational clauses distinct, we should be able to do away with any special assignment parameter and keep just the portion of models responsible for interpreting constants. And this is precisely what Wehmeier's theory does.

Technically speaking, the maneuver is flawless. But it also means that theories of this kind apply only to a very special class of languages – i.e., ones that have an infinite number of semantic primitives. Hence, they cannot be learned because they cannot be finitely presented. So it cannot even begin to account for how resource-bounded agents manage

to quickly and efficiently compute the meanings of arbitrarily complex expressions. On this front, the Tarskian theory fares much better. While it may initially look like it too employs an infinite number of semantic primitives in the form of the various denotations for each distinct variable, it is important to realize that the denotations are *uniform* in a way that the interpretation of an infinite sequence of constant symbols is not. By this, I mean that they all denote variants of the same operation (namely, projection). This particular form of uniformity ensures that the entire collection is easily representable in finite terms.[22]

At this point, the advocate of Wehmeier's proposal might complain that this is a misrepresentation of their theory. After all, they likewise see a high degree of uniformity among the meanings of constants. In fact, the entire family seems like it can be parametrically specified using the following basic recipe: the meaning of c is the function that maps a model $\mathcal{M}$ to whatever c denotes in $\mathcal{M}$. So all that an agent really needs to know in order to grasp all of these meanings is that models supply denotations to constants. This, of course, would be a pretty extreme departure from Frege's own views.[23] But more to the point, if it were really that easy to know the meaning of a constant, then it would, in a sense, be too easy to do any serious explanatory work. In particular, knowledge of Wehmeier's Fregean meanings could not account for the facts about an agent's linguistic competence that a compositional semantic theory is supposed to account for. After all, competence in the use of a name plausibly requires much more than *just* the general knowledge that it be used to denote

[22]One simple way of doing this runs as follows: all projections of an infinite sequence can easily be modeled by a single operation that removes the head of a list from its tail and returns each individual component separately. The first projection can then be identified with the result of applying this operation once, discarding the tail. And every subsequent projection is achieved by iteratively applying the same operation to the returned tail until the desired position in the sequence is reached. A potential language learner would then only need to acquire one very easy to specify operation, and have a general capacity to repeatedly apply the same operation to its own output (which is already assumed to be the case in basically *any* interesting language processing task).

[23]This approach has a conception of logical languages as formal objects that can admit many distinct interpretations at its heart. And this is certainly not how Frege saw his *Begriffsschrift*. See Van Heijenoort (1967) for a classic discussion.

something that it names.[24]

While I take the above considerations to be more or less decisive against this specific formal version of Fregeanism, we should not be too hasty in concluding that the general Fregean outlook is necessarily doomed to failure. After all, the objection hinges on somewhat specific features of a technical proposal, and we have no reason to think that every faithful attempt to formally articulate a Fregean theory is necessarily going to share those features. We would therefore like to find a more general objection that hinges only on features that we have good reason to expect to be present in faithful implementation of the core Fregean ideas. But where should we look?

By now, it should be clear that the fundamental differences between the Fregean and the Tarskian theories are best articulated in terms of the different grammars that they employ. According to the Tarskian grammar, formulas with free variables are genuine, fully interpretable constituents of quantificational claims. The Fregean grammar, on the other hand, needed to be carefully articulated so that it would never make reference to open expressions. But as we saw, this need to avoid the use of free variables also necessitated a leap upwards in terms of grammatical complexity. Since every Fregean theory is presumably going to try to avoid the use of free variables in a similar way, this increase in complexity looks like it is an unavoidable feature of any Fregean proposal. But this increase in complexity also looks like it has some problematic interpretive ramifications.

When a language is thought of as a collection of uninterpreted formal structures, it is easy enough to regard a grammar for it as an attempt at codifying a set of combinatorial

[24]The basic point being made here is sometimes given by advocates of the Davidsonian truth-definitional program in semantics as an argument against the Montagovian model-theoretic program (see, e.g., Lepore (1983)). The difference is that Davidsonians supply an unrelativized definition of truth for a language with a fixed interpretation, while model-theoretic semanticists supply relative definitions of truth where the interpretation of the language may vary. In practice, there is actually very little difference between the two perspectives since the relativized notion of truth does not appear to bear any serious explanatory burden in model-theoretic analyses of empirical phenomena (see, e.g., Glanzberg (2009)). Here, however, the problem is much more serious since the fact that the interpretation of the language is (in principle) variable plays an essential role in Wehmeier's semantics for condition abstraction.

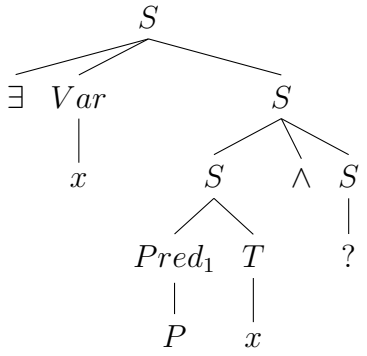
facts that are interesting and important in their own right. But this formal conception of language has very little contact with the reality that drives linguistic inquiry in the first place. Languages are ultimately only of interest to us because they are devices for efficiently conveying information, and this way of thinking about things requires that they be regarded essentially as interpreted structures. And from this perspective, it is very hard to get away from the view that the main (if not only) function of a grammar is to guide a procedure for interpretation. What, then, do the Fregean and Tarskian grammars suggest about the nature of this procedure?

Obviously, expressions do not wear their grammatical structures on their sleeves. The purely linear nature of a string hides whatever hierarchical organization the grammar endows it with. So whenever you are given an expression, some amount of computational work needs to be done in order to recover that hierarchical information. At this point, two possibilities for interpretation immediately suggest themselves – either interpretation waits for the input string to be fully processed, or it proceeds (as best it is able) to operate in tandem with the assignment of structure to the string. As a limiting case of the second option, interpretation is fully *incremental*. Each component of the string is read in a strictly left-to-right manner, and a partial interpretation is assigned to the portion of text that has been read so far.

Of course, the extent to which an incremental interpretation strategy even makes sense depends entirely on whether or not a theory has the semantic resources necessary to assign an accurate meaning estimate to any initial segment of a grammatical sentence. According to the Tarskian theory, at least, this is extremely straightforward. Suppose, for instance, that a language processor conforming to the Tarskian theory encounters the following initial segment of text:

$$\exists x(Px\wedge$$

The Tarskian grammar has no problem making sense of the organization of this string. Any incremental language processor would presumably (at least implicitly) assign it the following partial structure:



On the assumption that the consumed text is an initial segment of a fully grammatical unit, the processor knows that whatever it reads next will be a sentence-type constituent. Thus, it can use a function from sentence-type meanings to sentence-type meanings as its current estimate of the complete text's meaning. And which function that should be can be trivially read off the partial tree – in this case, it is just $\lambda\phi.[\![\exists]\!]([\![x]\!], ([\![\wedge]\!]([\![P]\!]([\![x]\!]), \phi)))$.[25]

The situation is much different in the Fregean case. As we just saw, the essential trick was to treat an incremental meaning estimate as a functional abstract whose type is fully determined by the expected types of the currently unread constituents that the assumption of grammaticality requires to be exemplified in the remaining incoming text. But these unread constituents may contain variables. And in general, the number and position of these variables cannot be safely inferred on the basis of the text preceding them. The Tarskian theory has no difficulty coping with this fact since each variable is an independently interpreted item and can be dealt with on an individual basis each time one is encountered. But the Fregean theory cannot similarly make sense of incoming variable-containing text since the grammar assumes that all of the variables bound by a particular quantifier are handled simultaneously. This means that whenever a quantifier is encountered, a Fregean language processor would need to essentially halt interpretation until its scope-target has been fully processed, at which point a correct condition could be abstracted and fed to the

[25]It is a more or less straightforward exercise to systematize these remarks in order to formally demonstrate that the Tarskian grammar is compatible with an incremental interpretation strategy. See Appendix A for details.

meaning of the quantifier.[26]

Viewed from the lens of a concrete language processing task, then, there is a clear difference that separates the Tarskian and Fregean theories – the Tarskian theory is compatible with incremental interpretation, and the Fregean theory is not. Whether or not this can be construed as an objection to the Fregean theory, of course, depends on what its purported explanatory aspirations actually are. If, among them, we find an attempt at explaining how human language users can grasp the truth-conditions of languages with quantificational devices like those of first-order logic, then facts about how human language processing actually proceeds are clearly relevant to evaluating its success.

For what its worth, human language processing at the sub-sentential level is now by-and-large agreed to be incremental to at least some degree.[27] For example, consider the following minimal pair:

(2) a. `The authors read in the garden were uninteresting.`

b. `The articles read in the garden were uninteresting.`

There is a readily appreciable difference between how difficult (2a) and (2b) are to interpret, despite the fact that they are structurally identical. In particular, (2a) gives rise to a very pronounced *garden-path* effect. As it is read in left-to-right order, `read` is first treated as though it was the main verb of the sentence. But this hypothesis must be discarded once the true main verb is ultimately encountered, which means that some amount of backtracking and re-analysis of the sentence is required to arrive at the correct interpretation. This effect is not nearly as strong in (2b). The obvious explanation, of course, is that `read` is never seriously hypothesized to be the main verb because that would result in an unreasonable

[26] A Fregean may protest and say that the correct thing to do is to make use of a family of abstraction operators $(-)_c[\xi]$ in the processor. In other words, the correct meaning estimate for an initial segment of text like '$\exists x(Px\wedge$' would be representable by a term like $\lambda\phi.[\![\exists]\!]((\![Pc]\!]\cap\phi)_c[\xi])$. But this clearly will not work since the correct abstraction operator to use clearly depends on features of the text that have yet to be encountered.

[27] See e.g. Haddock (1989) and the references therein.

reading in context – articles are not the kind of thing that can read, though they can be read. But this means that semantic and pragmatic considerations are being used to rule out competing syntactic analyses as they are being generated in real-time.[28] And if we start looking at linguistic units larger than sentences, the familiar kinds of examples from the literature on the exceptional binding abilities of indefinites suggest that quantificational dependencies can be carried on past clear interpretational boundaries:

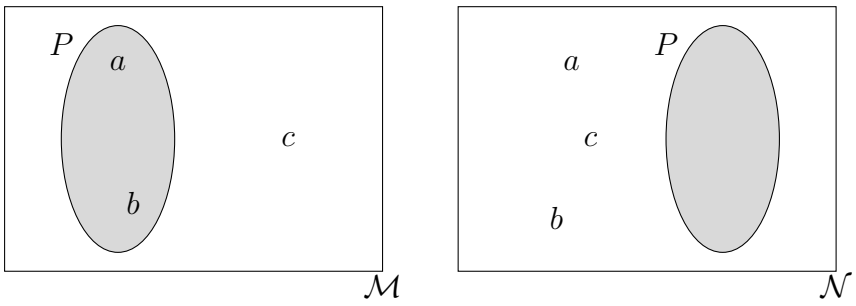
(3) a. `A man walks in the park. He whistles.`

b. `Kim plants a tree in every park she works on. It's always a pine.`

If we treat the indefinite in (3a) as something analogous to existential quantification, then it appears that we have a case where a quantifier extends its binding capabilities past a sentence boundary by controlling the pronoun in the following sentence. And (3b) lends more credence to the idea that the phenomenon in play here is genuinely quantificational, rather than explicable in terms of something like the referential indefinites posited by Fodor and Sag (1982). On its most natural interpretation, the indefinite in the first sentence of the sequence in (3b) takes scope under the universal quantifier, inducing a kind of functional dependency that is one of the main hallmarks of true existential quantification. In other words, the truth of an utterance of (3b) only requires is that it be possible to associate each park in question with a (possibly different) tree that Kim planted in it. The alternative reading (that Kim plants the same tree in every park) would be absurd. And notably, the pronoun in the subsequent sentence is naturally read as being controlled by that narrow-scoping indefinite. So it straightforwardly inherits that induced dependency and cannot be assigned a single tree as its content in context. Instead, it must be seen as being somehow bound by the indefinite, in much the same way that an existential quantifier binds a subsequent variable.

[28]In general, the ability to reason over incremental meaning estimates like this is actually one of the main advantages this kind of interpretation strategy has over its competitors. For example, in a language that exhibits a high degree of non-trivial syntactic ambiguity, these meaning estimates can be used as an extremely valuable mechanism for deciding between different competing parses as the sentence is processed in real-time. See e.g. Altmann and Steedman (1988).

If these examples are taken at face value, then it looks like a plausible theory of natural language quantification ideally ought to be compatible with a modestly incremental interpretation strategy. And while I happily admit these considerations really only directly bear on natural languages, there is no reason to suppose that they could not be generalized to bear on artificial cases like the one at hand. Artificial languages, after all, can also be understood and interpreted by human language users, and so there has to be an analogous set of facts about what this understanding actually consists in. And more directly, there do appear to be at least some facts about our grasp of first-order logic that are best explained by a theory compatible with an incremental interpretation strategy. In particular, we seem perfectly capable of making significant semantic determinations on the basis of only an initial segment of a first-order sentence. For instance, even though the string "$\exists x(Px\wedge$" is in a sense incomplete, it has obvious semantic implications – in particular, it can be used to distinguish between the following two models:



Given just this initial segment of text, we can definitively rule out $\mathcal{N}$ as a candidate model for the complete sentence, while $\mathcal{M}$ can still be maintained as a live option.

This shows that there is a class of readily understandable questions that relate proper initial segments of first-order sentences to certain classes of models. In other words, there appears to be a class of facts of a genuinely semantic character concerning the properties of initial segments of first-order sentences. Moreover, we seem to be able to quite easily determine the answers to at least some of these questions. A theory that is compatible with an incremental interpretation strategy has an easy time providing a semantic explanation of

these facts – nothing prevents us from straightforwardly interpreting these segments of text, and explaining their relationship to various classes of models in the ordinary way. Prima facie, then, it looks like a cognitively plausible theory of quantification should be compatible with an incremental interpretation strategy. And since the essential ingredients of a Fregean syntax look like they are fundamentally incompatible with this kind of interpretation mechanism, a Fregean theory will always fail whenever it is construed as being responsive to these kinds of processing considerations. The comparative simplicity of a Tarskian theory's grammar, on the other hand, renders it immune to this criticism.

At this point, however, a Fregean may protest that processing considerations of this kind are not relevant to the assessment of their proposal. According to this line of response, a semantics for a formal language merely aims to stipulate the relation between the formulas of that language and their meanings. Importantly, it does not claim that there will necessarily be a transparent relationship between the specification of that abstract semantic structure and its corresponding implementation in a concrete language processing system. While I will ultimately argue in the next chapter that the Fregean should endorse something like this idea, it is worth noting that this suggestion by itself is both more radical and less helpful than it might otherwise initially appear. As we have already noted, the standard expectation that a semantics for a language be compositional is normally motivated on cognitive grounds – a compositional semantics is supposed to explain (perhaps in a very weak sense) how it is possible for an agent with limited cognitive resources to nonetheless be in a position to interpret an unbounded number of novel complex expressions. So something like a compositional meaning assignment is normally assumed to be implicit in an agent's operative understanding of a language. This line of response, on the other hand, seems to suggest that the stipulated semantics given by the language designer has very little to do with the collection of computable heuristics people use to interpret it. And the Fregean must admit that we are still owed an explanation of these heuristics. But in this case, those heuristics may not even have any direct relationship to the grammar of the language itself.

So in breaking with the standard assumptions about the explanatory role that compositional meaning assignments play, the Fregean will have a much more difficult time specifying what those heuristics are and how they are related to their officially formulated semantic theory. So without any of these further details filled in, the Tarskian still appears to be in a much better position concerning this issue.

CHAPTER 2

Variables Revisited

The previous chapter's discussion has put us in a somewhat awkward position. Neither the Tarskian nor the Fregean accounts of quantification appear to be fully adequate, as each was seen to have its own strengths and weaknesses. The primary aim of this chapter is to relieve some of the apparent tension between these two perspectives. They may therefore be understood as complementing rather than conflicting with each other. And as I have already argued that the best way to appreciate the difference between the two perspectives is to reflect on the ramifications of their different grammars, it seems like the most reasonable way to proceed is by more directly scrutinizing the motivating intuitions that lie behind the design of those grammars.

In §1, I note that grammatical structure has been traditionally appealed to in two different kinds of explanatory enterprises. The first aims to explain the apparent productivity and systematicity of linguistic understanding. Grammar's primary role in this enterprise is to furnish structural descriptions of expressions suitable for effectively computing their meanings. The second aims to explain the notion of logical consequence in terms of the abstract formal structure of valid inferences. Here, grammar's primary role is to furnish structural descriptions of expressions that transparently display their so-called *logical forms* instead. I claim that these enterprises put different and sometimes even conflicting demands on structural descriptions, meaning that there is no general guarantee of finding a grammar for a given language that is equally well-suited to serve both explanatory roles.

In §2, I claim that the Fregean grammar performs much better than the Tarskian grammar

from the logical perspective. More specifically, I argue that a grammar for a first-order language that is well-suited to play the logical explanatory role will (among other things) need be capable of describing a structural relation that holds between a universally quantified sentence and each of its singular instantiations. And the Fregean grammar appears to be purposely designed in such a way that makes this possible. The Tarskian grammar, on the other hand, cannot give a similar description of this relation at a purely structural level of abstraction.

In §3, I claim that the situation is reversed from the computational perspective. In particular, the discussion from the previous chapter shows that the Tarskian grammar is much better than the Fregean grammar at providing reasonable inputs to a process of compositional interpretation. But this also points the way to a resolution between the Tarskian and Fregean perspectives. Since the two theories are apparently targeting different phenomena, there is no reason to think that they necessarily come into conflict with each other. If a language does not admit a single grammar suitable for both explanatory purposes, it will be necessary to think of the expressions in that language as having multiple structural analyses of quite different kinds. The Fregean and Tarskian grammars can then be understood as working in tandem, each filling a theoretical lacuna left by the other.

I conclude in §4 by returning to a discussion of the theoretical role of variables. I argue that because the Tarskian grammar is designed with the desiderata of the computational perspective in mind, it is more directly concerned with how humans actually understand quantified first-order statements. I suggest that the Tarskian understanding of quantification is instance-based, in the sense that quantified sentences are interpreted as statements about the presence or absence of suitable verifying witnesses. And this means that quantified claims turn out to be about the individuals who make up the domains from which witnesses may be drawn. In order for this kind of analysis of quantification to work in full generality, however, an agent will also need to track complex structural dependencies between different objects in the domain and so will need more than one way of thinking about those objects. I claim

that the role of variables in the Tarskian theory is to express guises of this general kind. This account then yields a straightforward diagnosis of Fine's antinomy of the variable.

2.1 Grammar in Logic and Semantics

We have already seen that the difference between the Fregean and Tarskian perspectives on quantification in part involves a difference in their respective grammars. We can get better insight into the inner workings of these two theories, then, if we more closely examine the reasons they each have for endorsing their grammar of choice. But in order to do this, we will first need to consider some of the general explanatory roles that grammar has been taken to play in logical and semantic inquiry. The Fregean and Tarskian grammars can then be understood as being tailored to serve these roles in one way or another. With this in mind, we may identify two roles that will be of particular interest – one *computational*, and the other *logical*.

The computational role of grammar ought to be extremely familiar from the previous chapter's discussion. A language can be abstractly identified with a set of interpreted expressions (expression/meaning pairs), and an agent who is proficient in that language has the potential ability to recognize and correctly use every expression in that set. But most non-trivial languages are infinitely large when viewed from this angle, and human agents only have a finite supply of cognitive resources. So proficiency with a given language has to be grounded on a finite base of some kind. The standard approach to this problem is to assume that humanly possible languages are minimally those that are somehow finitely generated. An agent's competence with a particular language will then be partly explicable on the assumption that they are in possession of a generating device capable of producing all and only its well-formed expressions.

In this context, grammars and generating devices are treated as being more or less interchangeable. This means that the possible grammatical structures assignable to an expression

are typically identified with the possible traces a generating device may leave when generating it. And since a given generating device is meant to partly model the linguistic capacities of language users, the properties of its corresponding grammar ought to be more or less in line with those capacities. In other words, the adequacy of a proposed grammar to serve the computational role is judged primarily in terms of its ability to explain certain facts about human language use in a psychologically plausible fashion. And while there is undoubtedly a healthy amount of disagreement among theorists about the exact range of facts that a proposed grammar is ultimately responsible for, most will now agree that at least some of them concern the apparent productivity and systematicity of linguistic understanding. Thus, compatibility with a plausible compositional interpretation is an extremely prized formal feature of a grammar since it is a means of explaining those facts too.[1]

This way of viewing grammar is most at home when some language is already common currency among a community of speakers. The task of the grammarian is then to systematize that language neatly, making the structural analysis of language forms a kind of descriptive enterprise. While this may be a perfectly reasonable attitude to adopt in the study of natural languages, it is unclear what bearing it has on formal languages like the one presently under investigation. As invented languages have stipulated grammars, the grammarian's task in that context is language design rather than language description. Here, how much the computational perspective on grammar makes sense depends entirely on the designed language's intended purpose. After all, many languages are designed with absolutely no regard for human learnability or ease of use because they are not something that humans are supposed to learn or use in the first place. In this situation, psychological considerations ought to play little to no role in specifying a grammar. But the language that we are presently considering is not like that. Frege originally presented his *begriffsschrift* as a refinement of the language of ordinary mathematical discourse that would be more suitable to certain scientific

[1]Note that commitment to the principle of compositionality as it is standardly formulated is from this perspective equivalent to the idea that generating devices generate interpreted rather than uninterpreted expressions. See Kracht (2011) for a systematic development of this point.

purposes. One of its primary expressed aims, then, was to facilitate the representation and communication of exact mathematical thought. So the computational perspective makes sense here, at least insofar as Frege intended his language to be actually learned and used by mathematicians and logicians and for their understanding of that language to be productive. And since there are no other antecedent facts about language use that need to be captured, it follows that this perspective would view the compatibility with a plausible compositional interpretation as the leading role of a grammar in this context.

The logical role of grammar, on the other hand, more directly concerns a very specific issue of language design. Its roots can be traced back to a particularly ancient idea concerning our intuitions about arguments like the following:

(4) a.
```
Some logician is a philosopher
Every philosopher is wise
∴ Some logician is wise
```

b.
```
Some regular polyhedron is a platonic solid
Every platonic solid is convex
∴ Some regular polyhedron is convex
```

c.
```
Some cat is a robot
Every robot is electric
∴ Some cat is electric
```

These arguments are noteworthy because they are all both safe and binding from an epistemic perspective. A reasoner incurs no risk of passing from truth to falsehood in moving from premises to conclusion and must endorse the conclusion if they also endorse the premises on pain of contradiction. This property of arguments is normally presented in broadly semantic terms. Each inference in (4) has these epistemic features because it is *valid*, in the sense that its conclusion cannot be false if its premises are both true. In this situation, what is of particular interest is the overwhelming intuition that they are valid for the same reason. In other words, the fact that the truth of their premises guarantees the truth of

their conclusion has nothing to do with their specific contents. Instead, it has something to do with a general form that the arguments in (4) all instantiate. This thought is normally presented by abstracting away from an argument's contentful parts:

(4′)
$$\begin{array}{l} \texttt{Some } A \texttt{ is } B \\ \texttt{Every } B \texttt{ is } C \\ \hline \therefore\ \texttt{Some } A \texttt{ is } C \end{array}$$

The residue of this abstraction is a schema that supposedly characterizes the purely logical structure of these arguments. The validity of general forms of this kind is then taken to be the real target of logical analysis. This project gives rise to another familiar way of thinking about sentence structure. Actually applying logic (conceived as the science of schematic validity) in any particular instance presupposes that we can identify which valid schematic patterns a given sentence is participating in. The so-called *logical form* of a sentence is a description of its structure that is relevant to judgments of this kind.

From the logical perspective, a grammar is adequate if it yields suitable structures for this purpose. Traditionally speaking, this property has been taken to be somewhat hard to come by. For instance, the history of the philosophy of language is replete with claims that the surface structures of natural language sentences are bad guides to their logical structures. Any grammar designed in part to capture those natural structures would thus be unsuitable for serving the logical role. Instead, it was usually assumed that great care had to be taken in crafting a specially designed formal language appropriate for this kind of purpose. From this angle, the logical analysis of natural language expressions became more of an exercise of translation into these specially designed languages rather than one of direct structural description. It is, therefore, somewhat misleading to talk as though the traditional notion of logical form is really a notion of grammatical structure, at least when natural languages are concerned. The best we could seemingly hope for was a grammar suitable for the computational role alone – the possibility of a grammar that could serve both roles equally well was entirely ruled out.

Even if we grant that the traditional line on logical form is correct, we may still wonder if a given formal language can have a grammar that is suitable for both roles. While they are not obviously in direct conflict with each other, they appear to concern very different structural properties. As we already briefly noted, when we apply the computational perspective to formal languages, it will be more or less exclusively concerned with securing the ease of computing the meanings of individual sentences on an ad hoc basis. It follows that a grammar suitable for this purpose will need to characterize the semantically relevant aspects of internal constituency. In other words, the grammar will need to provide the resources to accurately describe certain facts about how the components of an individual sentence are related to each other. The logical perspective, on the other hand, is primarily concerned with the structural classification of entire arguments. And since arguments typically involve more than one sentence, a grammar will need to make structural distinctions that will enable cross-sentential structural comparisons to serve this purpose well. More briefly, the computational perspective is concerned solely with *internal* grammatical properties, while the logical perspective is concerned partly with *external* grammatical properties, too.[2]

These different concerns will sometimes lead to competing judgments about structure. To take a particularly simple example, we may consider the structure of basic boolean combinations:

(5) a. A `and` B

b. A `or` B

From a computational perspective, (5a) and (5b) are grammatically isomorphic. Their semantically relevant internal structures can both be accurately described as the application of a propositional connective to the same two component subexpressions. They wind up meaning very different things because the concrete realization of that connective is different,

[2]While my presentation differs in many important respects, this basic observation is more or less the same one made by Iacona (2018).

and this is a lexical rather than structural difference. But given what we have said about the notion of logical form, (5a) and (5b) will obviously have to be distinguished since these two different connectives participate in two very different kinds of characteristic inference. Here is a very straightforward case in point:

(6) $$\frac{A \texttt{ and } B}{\therefore \quad A}$$

A notion of logical form has to be minimally fine-grained enough to discern that (5a) can participate in an inference schema like (6) and that (5b) cannot.

While these judgments are different, we have yet to see a reason to think that they are at cross-purposes with each other. For instance, while the logical perspective may lead to certain finer-grained distinctions than ones that can be motivated on the basis of meaning-combinatorics alone, it is not clear that a grammar that makes those additional distinctions would be necessarily unsuited to serve the computational role. An actual conflict will arise between them only if the logical perspective requires that certain sentences have computationally unreasonable internal constituency structures. And in what follows, I would like to suggest that when the grammar of quantification is viewed from these two lenses, it generates a conflict of exactly this kind.

2.2 Quantification and Logical Form

Frege was first and foremost a logician, so, unsurprisingly, the Fregean grammar was designed with many of the central desiderata of the logical perspective in mind. In order to make this point, we will need to first briefly recapitulate some of the basic linguistic insights that Frege brought to bear when he modernized the subject. The term logic of his processors was partly organized around the surface grammatical contrast between subject and predicate:

(7) a. `Socrates is mortal`

b. `Philosophers are mortal`

The surface constituent structures of (7a) and (7b) appear to be closely related. According to the traditional model of grammatical composition, this fact was taken somewhat seriously from a logical perspective. But as (7a) and (7b) logically behave in quite different ways, the result was a quite complicated picture of logical structure. According to a somewhat crude caricature of Aristotle's *De Interpretatione*, all simple propositions are arrived at by some operation joining subject to predicate. Subjects come in two varieties – individual and general – which is a distinction that is likewise inspired by surface grammar (roughly tracking the contrast between proper and common nouns). When the subject is an individual term, there are two ways of joining it to an accompanying predicate with differing semantic effects – either affirm or deny the predicate of the subject. When the subject is a general term, you may likewise affirm or deny the predicate of a subject. But as a general term may apply to more than one entity, affirmation and denial may be universal or particular. So here, there are instead four ways of joining it to an accompanying predicate which each correspond to one of the four different corners of the square of opposition. Hence, there are (at least) six different mechanisms for joining subject to predicate and forming a basic proposition, which are conventionally signaled by the use of auxiliary grammatical devices (e.g., syncategorematic expressions of various kinds). Schematically:

(8) a. a `is` B

b. a `is not` B

c. `Every` A `is` B

d. `No` A `is` B

e. `Some` A `is` B

f. `Some` A `is not` B

Inference patterns involving basic propositions so understood were then considered the principal target of logical analysis.

While many interesting and non-trivial results were obtained in this tradition, the emphasis on basic propositions also put some principled limitations on what could be fruitfully analyzed. A classical example of one of these limitations is the so-called problem of multiple generality:

(9)
```
A certain mathematician is admired by every philosopher
∴ Every philosopher admires one or more mathematicians
```

The above inference is obviously valid but involves claims whose quantificational structure is too complicated to be represented by any of Aristotle's basic propositional forms. A long-standing concern of logic was, therefore, to generalize the notion of logical form in a way that enables a plausible explanation of the validity of arguments like (9).

One of Frege's most noteworthy logical accomplishments was the simplification and systematization of the logic of quantification along those lines. In order to do this, he first reformulated the grammar of certain fundamental logical notions using expressive idioms borrowed from mathematical languages rather than natural ones. As we have already seen, he first replaced the traditional subject/predicate contrast with the function/argument contrast. This allowed him to reduce the basic modes of propositional composition from the traditional six to the single particularly simple and well-understood process of saturating a function with its required argument(s). As logical particles were previously treated as syncategorematic markers of different basic modes of composition, this simplification meant that he would instead have to give categorematic treatments of them as functions of some kind. The negation and conditional operators were given their familiar truth-functional treatments, and the universal quantifier was treated as a higher-order function which, when combined with the characteristic function of a set, would yield the True just in case that set coincided with the universe. But while the universal quantifier was given this categorematic interpretation, Frege treated its grammar somewhat differently than other function-denoting expressions in his language. In particular, he borrowed another idea from mathematical languages and pre-

sented it as a *variable-binding* operator.[3] It thus additionally involved introducing variables that were used to mark the positions in its syntactic argument that it controlled and were used to determine the property it needed to take as its logical argument. This move granted Frege's language enough expressive flexibility to give an extremely elegant solution to the problem of multiple generality:

$$(9') \quad \frac{\exists x(Mx \wedge \forall y(Py \rightarrow Ayx))}{\therefore \quad \forall y(Py \rightarrow \exists x(Mx \wedge Ayx))}$$

The translation in (9′) is the modern characterization of the logical structure of the argument in (9) that Frege's logical grammar enables. So his language looks like it has all of the expressive resources it needs to represent traditionally difficult patterns of quantification. All that he needed to do was supply the basic resources required to explicate the validity of these patterns. In order to do this, Frege supplied a system of logical axioms and rules of inference. The idea was to justify the logical truth of a handful of axioms directly and ensure that the rules of inference preserve logical truth. We can then safely derive yet further logical truths by chaining together axioms and inferences. In this context, we can say that a conclusion C is a logical consequence of a set of premises $P_1, \ldots, P_n$ when the conditional $(P_1 \wedge \ldots \wedge P_n) \rightarrow C$ is derivable in Frege's system.[4] This system was extremely powerful and certainly sufficed to demonstrate the legitimacy of (9′).

For our present purposes, we can focus on one small aspect of the way that this system handled universal quantification:

---

[3]More precisely, he seemed to be partly inspired by the already quite pervasive mathematical practice of using variables to express general truths (e.g., $(x + y)^2 = x^2 + 2xy + y^2$). But this only allows for the expression of widest-scope universal quantification. So he adopted the idiom of variable binding operators that can occur sentence-internally (e.g. the occurrence of $\lim_{x\to\infty}$ in the specific claim $\lim_{x\to\infty} \frac{1}{x} = 0$) to mark internal scope boundaries.

[4]Frege's emphasis on the notion of logical truth rather than schematic validity and consequence is a little bit out of step with many of his forebears. His own attitude seemed to be that these two perspectives were more or less interchangeable (see, e.g., his remarks in Frege (1884), §17).

$$(10) \quad \frac{\forall x \phi(x)}{\therefore \quad \phi(t)}$$

The above schema captures one of the most fundamental principles governing the logical behavior of the universal quantifier – that what is true of everything is true of anything. This inference was given pride of place in Frege's logic, as it was taken (in sentential form) as one of his axioms. Given the central position that this inference pattern occupies in the Fregean framework, a grammar capable of adequately serving the logical role will minimally need to be able to structurally discern when a pair of sentences is participating in it. In other words, the grammar will need to be able to describe the structural relation that holds between a universally quantified sentence and each of its singular instantiations.

The Fregean grammar is capable of immediately delivering this kind of result. According to the Fregean, the internal structure of a universally quantified sentence like $\forall x \phi(x)$ is really the result of combining the (second-level-function-denoting) universal quantifier $\forall$ with the (first-level-function-denoting) condition $\phi[\xi]$. And that condition is, in turn, syntactically derivable from any of its substitution instances. Thus, while it is presumably incorrect to say that $\phi(t)$ is a syntactic constituent of $\forall x \phi(x)$ according to the Fregean, something very close to that claim is right. More specifically, for any choice of term t, the Fregean grammar posits a possible derivational history of the expression $\forall x \phi(x)$ that runs through the expression $\phi(t)$.[5] The grammar thus ensures that there is a deep structural relation between the premise and conclusion of any instance of the core schema in (10).

This is a clear and obvious advantage that the Fregean grammar has over its Tarskian counterpart. According to that proposal, the internal structure of $\forall x \phi(x)$ results from combining the quantifier phrase $\forall x$ with the open formula $\phi(x)$. But here, the result of uniformly substituting t for x in $\phi(x)$ bears no privileged structural relation to $\forall x \phi(x)$ that is not also

[5]Note that this formal property more or less coincides with the intuitive notion of syntactic constituency when a grammar is context-free. The fact that the Fregean language of predicate logic is not context-free means that these notions may come apart if constituency is defined in some other way.

shared by some illegitimate substitution instances, as well. For instance, *Raa* and *Rbb* both follow from $\forall xRxx$ while *Rab* does not. Therefore, a decent notion of logical form will need to specify a structural relation shared by the pairs $\langle \forall xRxx, Raa \rangle$ and $\langle \forall xRxx, Rbb \rangle$ that is lacked by $\langle \forall xRxx, Rab \rangle$. As we just saw, the Fregean accomplishes this utilizing a grammatically grounded notion of (quasi-)constituent-hood. But the Tarskian has no similar avenue of escape since *Rxx*, *Raa*, *Rbb*, and *Rab* are all derivationally independent of each other. In some contexts, this problem is explicitly addressed by distinguishing between two distinct notions of subformula. On this way of doing things, $\phi(t)$ as stipulated as being an (extended) subformula of $\forall x\phi(x)$ for all terms t, while the only (literal) subformula of that form that $\forall x\phi(x)$ has is $\phi(x)$.[6] But importantly, the extended notion of a subformula has to be understood as a *supplement* to the grammar rather than something it can independently deliver.[7]

2.3 Quantification and Interpretation

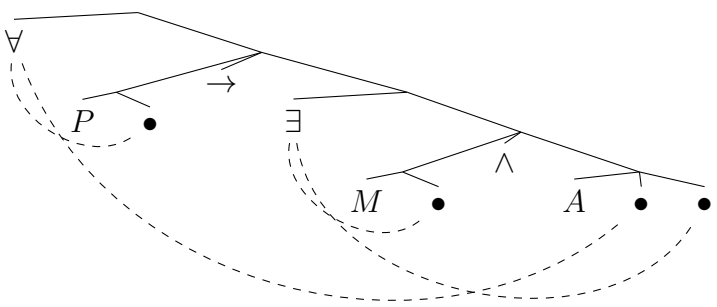
Unfortunately, this advantage of the Fregean grammar is a double-edged sword. While it can structurally characterize a useful notion of (quasi-)constituenthood that suits the logical perspective, it cannot characterize a plausible notion of (true) constituenthood that suits the computational perspective. In other words, the structures that the Fregean grammar generates are not plausible inputs to a process of efficient compositional interpretation. The basic point is best appreciated with a simple example:

(11) a. $\forall y(Py \rightarrow \exists x(Mx \wedge Ayx))$

[6]See e.g. the contrast between *Gentzen* and *literal* subformulas in Troelstra and Schwichtenberg (2000).

[7]The conceptual importance of (quasi-)constituent relations holding between the premises and conclusion of basic patterns of inference is frequently emphasized in the literature. For a classical discussion, see Hacking (1979). Another way to make the same basic point that was made in this section is just to note that if the subformula property really does have the kind of logical significance it is sometimes claimed to have, a logically adequate grammar ought to be able to structurally make sense of the relevant notion of subformula required for its statement. As we have just seen, the Fregean grammar has no problems here, while the Tarskian grammar will need to be augmented somehow.

b.



As we have already made clear, a highly significant aspect of the Fregean language is its ability to efficiently represent the logical structure of sentences involving more than one quantifier. And the Fregean grammar endows these sentences with two distinct kinds of structure – a hierarchical constituent structure (conventionally signaled by the use of parentheses) and a lateral quantifier-control structure (conventionally signaled by the use of variables). So the Fregean holds that the logical structure of (11a) is faithfully depicted by the diagram in (11b). But these two forces directly work against each other in this case – the lateral quantifier control aspect of structure foils the ability for the natural notion of constituent to do plausible work that is faithful to the other core tenets of Fregean semantics. To see the problem, remember that the Fregean takes the semantic contribution of a quantifier to be a second-level function that needs to be saturated by a first-level function of one argument. And at the topmost hierarchical level of (11b), constituenthood nicely coincides with this semantic requirement. The quantifier expression $\forall$ locally combines with the condition expression $(P\xi \rightarrow \exists x(Mx \wedge A\xi x))$, which has as its interpretation the (monadic) property of being something such that if it is a philosopher, then it admires a mathematician. But note that this is the *only* locally interpretable configuration that this sentence has. The interpretive coherence of the lower hierarchical elements essentially depends on the presence of the topmost quantifier filling some of their open positions.

This general problem can be put into sharper focus by considering asking what monadic property the inner existential quantifier of (11a) is locally combining with. Since we are

treating the existential quantifier as a higher-order property, we have to give it scope over whatever it takes as its argument. But now, we appear to have to do the impossible by giving it an argument that paradoxically involves elements (in this case, the occurrences of the variable x) functionally dependent on a hierarchically superior quantifier. The issue that we are now dealing with is the pervasive *non-locality* of quantification in first-order logic. A quantifier can semantically control syntactic positions that are arbitrarily far away from where it actually appears in the expression. This by itself is not an issue, but the Fregean additionally enforces the extra constraint that every quantifier's local compositional argument be a monadic property. And there is no good notion of monadic property that could possibly serve as the interpretation of an inner quantifier's local syntactic partner that *also* incorporates the capacity to eventually be partially controlled by yet more distant quantifiers. But note that this bit of reasoning presupposes that there is a *unique* answer to the question that we are asking. The natural Fregean response is to simply deny this presupposition.

According to the Fregean grammar, the positions that distant quantifiers govern are initially controlled by terms that are then abstracted away on-demand as the governing quantifiers are introduced in a given syntactic derivation. This means that the local syntactic argument of the inner quantifier in (11a) depends on the term that was abstracted in the formation of the condition expression $(P\xi \rightarrow \exists x(Mx \wedge A\xi x))$. For instance, if this expression were derived from the closed instance $(Pb \rightarrow \exists x(Mx \wedge Abx))$, then the local syntactic argument of $\exists$ would have been the condition expression $M\xi \wedge Ab\xi$. This means that in a syntactic *derivation*, at least, each quantifier gets locally paired with an expression that plausibly designates a monadic property. One way that a Fregean could proceed, then, would be to take the notion of a syntactic derivation as the real level of structural description relevant for compositional semantic analysis.[8] There would not be a unique semantic argument of the inner quantifier in (11a), as the answer to that question essentially

[8]Note that this is exactly how Wehmeier (2018)'s formal implementation of the Fregean proposal worked.

depends on how the sentence was derived, and it is in principle compatible with infinitely many distinct derivations. Of course, this move means introducing a considerable amount of spurious ambiguity – sentences like (11a) would potentially have infinitely many different parses that make no difference to the overall result of compositional interpretation. And that would undermine the idea that the Fregean grammar generates plausible logical forms since the sentences in a specially designed logical language ought to have only *one* possible logical form. This kind of proposal thus does not avoid the need to make a strict distinction between the logical and semantic structures that get assigned to an expression.

A Fregean might try to avoid making this distinction by instead holding that the meaning of a condition expression is derived from all of its possible substitution instances rather than an arbitrarily selected one. The inner quantifier of (11b) would not have a unique semantic argument because *every* condition of the form $(Mc \wedge A\xi c)$ needed to be considered in the computation of the meaning of the larger condition the quantifier in question is part of. This would involve endorsing a partially holistic picture of semantic interpretation. While the meaning of a quantified sentence may not depend on its relation to *all* of the other expressions in the language, it would depend on a *much* wider swath of them than is generally assumed. This way of thinking has many conceptual advantages. For instance, most discussions of Fregean property abstraction typically assume that it is a non-holistic operation – the meaning of a condition needs to be derivable from any of its substitution instances. And that leads to many thorny technical issues.[9] But many of these issues can be avoided if a property is instead abstracted as the residue of what is common to the entire class of the meanings of a condition expression's possible substitution instances.[10] Unfortunately, this way of thinking is also clearly at odds with the computational perspective's theoretical desiderata. That viewpoint is principally concerned with how to compute the meanings of

[9]See Humberstone (2000) for an interesting discussion

[10]Note that this way of thinking also coheres much better with how we normally think about the process of abstraction in other contexts (e.g., our grasp of the number **3** is arrived at by appreciating what is common to every three-element collection)

sentences on a reasonably ad hoc basis, in relative isolation from one another. And it would be completely unfeasible to have to compute the meanings of infinitely many substitution instances first just to compute the meaning of a single abstracted condition.[11] So this kind of Fregean would still ultimately have to offer a story about the process of semantic interpretation that is primarily driven by something other than the official structure that their grammar assigns to expressions being interpreted.

If the above reasoning is correct, the Fregean will need to supplement their proposal with an assignment of a compositionally interpretable structure to every well-formed expression. The lazy choice, of course, is just to use the syntactic derivations produced by the individual runs of their grammar. While these are not the true grammatical structures of the generated strings, they are still more or less immediately delivered by the theory and will do the job so long as the needed abstraction operator is reasonably definable. Of course, this involves a significant promissory note. And as we saw in the previous chapter, these structures additionally have undesirable computational properties that make them somewhat ill-suited to serve the computational role as well as other proposals. The alternative, then, is to use a class of structures specially designed for this purpose and are therefore more computationally well-behaved. The price we pay for this option is that more theoretical work needs to be done to relate the officially generated structures to the newly designed ones. But the explanatory gains we get may offset this cost. My suggestion is that the Fregean can regard the Tarskian theory in exactly that way.[12]

The Tarskian proposal is explicitly designed to compositionally deliver the *truth-conditions* of first-order formulas – after all, its most frequently encountered form is just a straightfor-

[11] Also see the traditional arguments against semantic holism that invoke considerations of learnability or compositionality in some form, like the one presented in Dummett (1991) or those cataloged in Fodor and Lepore (1992).

[12] Note that historically speaking, the Tarskian theory also seems to fit our earlier characterization of the computational explanatory enterprise better than you might guess. After all, Tarski was arguably engaged in an exercise of descriptive grammar, as the languages he semantically analyzed had already been common currency among working logicians for decades. His theory thus had to deliver the truth conditions that the expressions were already understood to have in an (admittedly small) community of established speakers.

ward definition of truth. And if this is the *only* thing that the Tarskian is trying to do, then compatibility with the Fregean proposal will be ensured so long as the truth-conditions it yields for closed formulas are the same as the Fregean ones. In order to design syntactic structures that can better serve this important role in interpretation, it needs radically to reconceptualize the grammar of quantification first. What was once a single syntactic unit in the Fregean theory gets decomposed into two distinct pieces – a sentential operator used to mark the scope and force properties of a quantifier and a variable that allows a quantifier to locally enter a syntactic derivation as though it were a constant. In making this distinction, the essential syntactic connection in the Fregean theory between a quantifier and the distant positions it controls is completely lost. First-order expressions are made to be grammatically well behaved at the cost of a transparent representation of their true logical form. But the Tarskian need not be concerned about this loss since capturing *that* structural relationship was never one of their goals. Their aim, rather, is just to fill a theoretical lacuna left by the Fregean account, and for that kind of task, they are more or less free to specify any class of structures that best serves that role. So long as we are careful not to conflate the two theoretical purposes a grammar could play in a theory, there is no reason to see any tension between the Fregean and Tarskian perspectives.[13]

2.4 The Theoretical Role of Variables

While the Tarskian and Fregean agree on the truth-conditions of first-order sentences, the fact that they analyze the grammar of the underlying language differently at least partly

[13]The idea that a grammar that is good for describing certain important structural regularities may be different from one that is good for describing the workings of a concrete processor is not new. In linguistics, the so-called *strong competence hypothesis* of Bresnan and Kaplan (1982) roughly states that the cognitive significance of the competence grammars typical of those given in the Chomskian syntactic tradition amounts to their being the same grammars implicit in actual human language processors. Merely granting that this hypothesis could turn out to be false requires that we be sensitive to this potential grammatical distinction. My suggestion is that sensitivity to this distinction can also be put to good work in philosophical contexts and that first-order logic provides a clear example of a situation where we should naturally think that the two kinds of grammar come apart.

motivates the idea that they differ with respect to a finer notion of content. After all, our judgments about the content of an expression often go hand in hand with our judgments about its subject matter, and those, in turn, are generally assumed to be directly reflected by grammar. For instance, take the classical dispute between Russell (1905) and Strawson (1950):

(12) `The author of "Über Sinn Und Bedeutung" was wise`

Strawson took the structure of the proposition expressed by (12) as being in line with the apparent surface structure of the sentence itself. He thus held the position that (12) was a simple and straightforward predication, and so about a particular individual. Russell, on the other hand, famously took the structure of the proposition expressed by (12) to have a quantificational character instead, which was, in turn, understood along broadly Fregean lines. Taking a few expository liberties in our presentation of Russell's view, we can say that he now essentially thought that the definite article `the` designated a certain higher-order relation which holds between two properties just in case they overlap and the first is uniquely instantiated. And if that is right, then (12) is really a kind of ternary relational configuration in which that higher-order relation is said to hold between the properties of being an author of "Über Sinn Und Bedeutung" and being wise – a claim that does not directly concern individuals at all. Their dispute was thus at least to some extent about the real subject matter of sentences like (12), and both were inclined to elucidate their respective positions in partly structural terms.

The differences between the Fregean and Tarskian grammars likewise motivate a kind of divergence in subject matter. A simple example suffices to make this point:

(13) $\exists x Px$

The Fregean grammar understands (13) as having the form of a higher-order predication. The statement it is making is *about* the property indicated by P and a feature that it allegedly enjoys (non-emptiness). The Tarskian grammar, on the other hand, cannot directly support

that kind of story.[14] It takes the open formula Px to be an independently meaningful constituent of (13). And since open formulas are not semantically distinguished from their closed counterparts, the interpretation of that formula will seemingly have to be something akin to a proposition rather than a monadic property. So the Tarskian is most naturally understood as taking (13) to really concern the modification of something (quasi-)propositional. While it is not immediately apparent how to best further elaborate on this idea, I suggested in the previous chapter that the Tarskian (subject-matter-individuated) content of (13) could be glossed in broadly modal terms. Thus, (13) says of the variable x that it is possible for it to assume a value such that Px is true. In other words, it is a claim about the possible values of x – namely, that at least one of them has the property indicated by P. So (unlike the Fregean), the Tarskian takes quantified claims to more directly concern the individuals that constitute the domain of quantification, rather than (just) the properties defined over that domain.

If we resolve some of the apparent tension between the Fregean and Tarskian in the way that I have suggested, the fact that their respective pictures diverge on this point is not a serious problem. The relevant notion of fine-grained content at issue does not seem to play any serious role in the development of logical or semantic theory. So neither the logical nor computational perspectives on grammar will directly enforce a constraint that requires an account of structure to yield specific judgments about subject matter. Instead, it seems better from our perspective to say that both ways of talking are legitimate. But the fact that the Tarskian and Fregean grammars are offered from different theoretical vantages makes the resulting content elucidations appropriate at different levels of abstraction. The Fregean grammar was concerned with characterizing a very high-level notion of logical form that respects the structure of an independently motivated (and mind-independent) type-theoretic

[14]Of course, it is a very popular strategy to talk as though it could. But this involves postulating a very stark ambiguity between free and bound occurrences of variables, and that clearly misunderstands how variables function in the Tarskian framework. It assigns variables a univocal interpretation, which can then account for their two different kinds of use. Ambiguity is the wrong diagnosis.

universe (remember that it takes the saturation of a function by its required argument was the main driving force behind syntactic as well as semantic combination). The Tarskian grammar, on the other hand, was claimed to be more directly concerned with lower-level issues of concrete processing and human understanding. It is thus not *merely* concerned with the world but also with how an interpreter cognizes it.

This last observation actually sheds much light on the conceptual underpinnings of variables. The point is best appreciated by more closely examining the semantic role that variables play in the Tarskian theory (now seen through a computational lens). Human competence with quantification undoubtedly consists of a complex and varied set of capacities. And some of the most centrally important of those capacities concern how we are inclined to assess the truth of quantificational claims. That is to say, part of understanding what claim a sentence involving one or more quantifiers is making involves (at least implicitly) recognizing one or more processes that would suffice for either verifying or falsifying that claim. As the Tarskian theory explicitly adopts the computational perspective on grammar, it can be thought of as trying to shed light on at least some of these capacities. Variables play an essential part in how this is going to work. A simple example will help illustrate what I mean by this. A first-order formula like (13) can be associated with a certain kind of abstract verification procedure, which can be roughly paraphrased as follows: first, empty the memory register labeled by x, then non-deterministically insert a value from the domain into that register, and finally check if that value is in the extension of P. This procedure succeeds if the answer is affirmative and fails otherwise. Its truth-conditions can then be defined in terms of this procedure – we say that it is true just in case there exists at least one successful run of the procedure.[15] Note that on this way of setting things up, the role that variables play in the theory is to label an abstract cell in an idealized machine's memory. Therefore, a variable can be understood as something like an address of a container for

[15]The treatment of existential quantification as non-deterministic success was first formally suggested in the context of first-order dynamic logic (see, e.g., Harel et al. (1977)) and was later famously adapted by Groenendijk and Stokhof (1991) in the design of their Dynamic Predicate Logic.

storing values taken from an underlying domain of quantification. The semantic role that a variable plays can then be understood in terms of the properties of those containers. But importantly, every container so understood is the same in the sense that they all serve that same abstract functional role. The resulting view is one in which variables are only *semantically* distinguishable in terms of the modal properties of the containers they designate – i.e., the brute fact that their contents *may* be distinct, even though they might not *actually* be distinct.

The story that the Tarskian just gave establishes the idea that human understanding of first-order formulas like (13) is partly instance-based, in the sense that they are interpreted as statements about the presence or absence of suitable verifying witnesses.[16] So (as we have already noted), quantified claims turn out to in part concern the individuals that make up the domains those witnesses may be drawn from. But if this is right, we will also need to recognize the necessity of a perspectival element in the semantics of quantification. In particular, a cognizer will need several distinct ways of thinking about those individuals:

(14) $\forall x \exists y Rxy$

A witness to a claim like (14) is (famously) a way of pairing each element of the domain with one or more partners that it bears the indicated relation to. So understanding a claim like this seems to naïvely require that an agent is in a position to think about the same domain under two separate guises – crudely, as *the potential subjects of the relation indicated by* R and *the potential objects of the relation indicated by* R. Having multiple guises of this general form then lets an interpreter track the complex patterns of dependence that may be required to verify a quantified sentence. And in general, the more distinct variables that sentence has, the more distinct ways of thinking about the same domain of quantification the Tarskian theory will need to assume are at an interpreter's fingertips.

[16]This is then borne out by the observation that people will, in fact, attempt to verify an existential claim by looking for a witness (or equivalently, falsify a universal claim by looking for a counter-instance). This way of thinking about this is consonant with recent empirical work on certain cognitive aspects of quantification. See e.g. Pietroski et al. (2009) and Szymanik (2016).

We can perhaps better appreciate the need to recognize distinct guises of this kind when we consider the role that variables play outside the insulated special context of the semantics of first-order languages. Consider, for instance, how they are used in ordinary mathematical discourse. One very prominent kind of use is in proofs like the following:

> **Claim.** There are infinitely many prime numbers.
>
> *Proof.* Let $\{p_1, \ldots, p_m\}$ be a finite collection of primes. Consider $n = p_1 \cdots p_m + 1$. Then just note that n has some prime divisor p_k, and that $p_k \notin \{p_1, \ldots, p_m\}$, since that would imply that p_k divides both $p_1 \cdots p_m$ and n, and hence also their difference $n - p_1 \cdots p_m = 1$, which is impossible. □

The surface structure of this proof plainly involves variables of several different sorts, which we will assume is a fact that ought to be taken seriously. Some of these variables are obviously associated with distinct ranges. For instance, p_k is stipulated to range over primes, while some of the numbers that n ranges over are composite. So different variables are sometimes straightforwardly just devices for talking about different things. That might tempt someone into thinking that the ranges of variables are all that is semantically relevant in the interpretation of the texts that contain them. But it is important to realize that in this context, at least, it is the way that the variable picks out its range that is actually semantically relevant. As a further example, note that while (say) p_1 and p_k range over the same things, they do so in such a way that ensures that their values are always distinct – p_1 and p_k both *may* designate any prime number, but they *must* designate distinct ones. Then just note that this proof cannot concern any one particular pattern of values assumed by the variables in question. Instead, the legitimacy of the argument requires that we consider *all* of the possible values the variables involved are allowed to assume simultaneously. And part of appreciating that fact requires that we understand that p_1 and p_k never assume the same value. In other words, part of interpreting this proof requires attending to and appreciating several distinct ways of talking about the same kind of thing.

Given this observation, it is natural to identify the Tarskian interpretations of variables with placeholders for the requisite guises of this kind. In order to handle quantified formulas of arbitrary complexity, it assumes that an agent has infinitely many of these, with no conventional assumptions about their antecedent structural relationships to each other.[17] But it avoids making any further assumptions about their character beyond the mere fact that they are somehow available to an interpreter. They are thus devoid of any conventional descriptive content that guises of this kind are normally associated with. All that's left to be said about the interpretation of x is that it is a way of talking about the things labeled by x.

We have now arrived at the idea that variables express concepts for thinking about potentially complex patterns of plural designation. In other words, each variable expresses a distinct way of thinking about the objects in a domain of quantification whose values may or may not be structurally related to the values of other variables. While this may sound strange at first, it is important to realize that nothing is particularly exotic about this idea. Ordinary thought and discourse seems to contain devices of this kind already. As an illustrative example, imagine that the participants in a seminar are each required to present and defend the contents of a different paper. Additionally, they are also required to provide critical remarks on one of the papers that their colleagues presented on. Finally, further suppose that no two students are allowed to comment on the same paper, so that each paper has precisely one presenter and precisely one commentator. In this situation, it seems that we can talk about the same collection of papers in three quite different ways: as *the papers discussed in the seminar*, as *the papers presented by the seminar participants*, and as *the paper criticized by the seminar participants*. And these three different ways of thinking about the same group of papers strike us as cognitively distinct, even though they may even

[17] Compare this situation with the semantics of Wehmeier (2018) that was discussed in the previous chapter. That theory also assumed that an interpreter had infinitely many distinct guises for thinking about objects in the domain (in the form of an infinite sequence of interpreted constants). Something like this conclusion thus seems hard to formally avoid in giving a compositional semantics for first-order languages.

be necessarily co-extensive in this context (assuming the rules of the assignment are not allowed to vary from world to world). This cognitive difference, therefore, does not seem to obviously reside in their modal profile but rather with the implicit structural relations they are assumed to bear to the domain of students. That is to say, thinking about the papers as *the ones discussed in the seminar* is structurally neutral, while (for example) thinking about the papers as *the ones criticized by the seminar participants* structurally relates them to a domain of criticizers.

If what we have been saying is right, then we are now in a position to give a perfectly straightforward and reasonable response to Fine's supposed antinomy of the variable. Recall that in the previous chapter, this puzzle was presented as one of the main reasons for abandoning the Tarskian semantic framework in favor of something more complex. It was motivated by the observation that distinct variables like x and y appeared to have exactly the same semantic role, while distinct pairs of variables like $\langle x, x\rangle$, and $\langle x, y\rangle$ appeared to have different semantic roles. This then generated a dilemma for any semantics of a language in which variables were treated as genuine lexical items – either the semantic values for the variables have to make more distinctions than those that can be motivated on the basis of semantic role alone (and hence are not really 'semantic' in some sense), or the semantics of the language cannot be compositional. But now we can see that the driving intuitions motivating the puzzle arise from a failure to respect a very modest and sound piece of methodological advice – namely that in order to determine the meaning of an individual lexical item, you should consider how it behaves when it is embedded in a larger expression.[18] For instance, consider the difference in meaning between definite descriptions like `the president of the United States` and `the current president of the United States`. Because these two descriptions can be used to talk about the same individuals, a careless theorist might be

[18]This is obviously related to Frege's so-called context principle, which urges us to never "ask for the meaning of a word in isolation, but only in the context of a sentence" (Frege (1884)). The piece of methodological advice presented here is an *extremely* weak reading of what Frege had in mind – presumably too weak to be correct as a presentation of Frege's own views.

tempted to confuse denotation with semantic role and say that they mean the same thing. But this would neglect the fact that they behave in manifestly different ways when they are embedded larger environments:

(15) a. `The president of the United States was a republican in 2018`

b. `The current president of the United States was a republican in 2018`

It is clear that (15a) can mean something different than (15b) – if both were uttered at the time of writing, then (15a) would be naturally interpreted as saying something true while (15b) would be naturally interpreted as saying something false. So these two descriptions must mean different things, even though they are both ways of talking about the same people.

The temptation to identify the semantic roles of x and y is entirely analogous to the temptation to identify the meanings of `the president of the united states` and `the current president of the united states`. A general fact about their use (that both may be used to talk about the same things, assuming that their ranges are the same) is being taken as a good enough reason to identify their semantic contributions. But we now know that this is clearly a mistake, as it appears to be once again confusing semantic role with denotation. From our perspective, the interpretations of x and y are different ways of thinking about the same domain. And since these different ways of thinking are required for tracking the complex patterns of dependence that may emerge over the course of evaluating a quantified expression, there will be broader contexts that x and y can be embedded in that will serve to distinguish them. The fact that the pairs $\langle x, x\rangle$ and $\langle x, y\rangle$ have clearly distinct semantic roles is precisely a case in point. Two occurrences of the same variable must always assume the same value, while two occurrences of distinct variables may assume distinct values. So if you use the same variable twice, you are in effect tracking a different pattern of dependence than you would otherwise be if you had used two different variables instead. In other words, the fact that distinct variables are different ways of talking about

the same things semantically matters in this kind of situation. So long as we remember that Tarskian variables are to be understood intensionally, the temptation to identify the semantic roles of distinct variables can be avoided.[19]

[19]The worst that the antinomy establishes is that the *fixed* association of a concept to a label that the Tarskian framework employs is somehow inappropriate because the labels x and y can never be used in different contexts to talk about the domain of quantification in the same way. Whether or not this is actually a problem is unclear to me, but it can easily be remedied by assigning variable labels a character which in context selects for the first unused suitable concept from some fixed enumeration. On this way of doing this, the formulas Rxy and Ryx turn out to mean the same thing (roughly that *the first ones designated* bear the relation indicated by R to *the second ones designated*). See Pickel and Rabern (2016) and Dever et al. (2018) for some examples of how to formally develop this suggestion.

CHAPTER 3

The Role of Anaphora in Interpretation

Our concern up until this point has been with variables in first-order formal languages and (to a much smaller degree) dialects of mathematically augmented English. And I believe that we have managed to identify a plausible semantic role that they play in those contexts. But that still leaves open an important question – namely, are there things that play the role of variables in (non-augmented) natural languages, too? Someone may, of course, be naïvely skeptical of this claim. After all, natural languages certainly lack anything like the overt variables that abound in the languages we have considered so far. But this would be a mistake since we now understand variables along broadly semantic lines – namely, as expressions that designate concepts for thinking about a domain of objects in a potentially structured way. And I have already suggested in passing that natural languages contain many devices for expressing concepts of this kind. So it now seems like a good time to turn to a more direct and sustained argument for that claim.

In order to accomplish this task in a convincing way, we need to argue that there are certain expressions in natural language which are both independently interpretable and theoretically required to track patterns of structural dependence between the elements of one or more domains of quantification. I want to suggest that pronouns are precisely what we are looking for. In many ways, this proposal should not be all that surprising. Pronouns are frequently analogized with variables because they also can appear both by themselves or semantically controlled by a commanding quantifier phrase. But despite this apparent formal analogy, there is very little agreement about what this actually says about the proper se-

mantic treatment of pronouns. My hope is that our present understanding of the theoretical role that variables play can shed more light on this problem.

This task will understandably take some time to complete, so this chapter's aim will have to be quite a bit more modest. For reasons that will hopefully become clear later on, the account of pronouns that I would like to present gives pride of place to their anaphoric occurrences. This chapter serves just to introduce some basic facts about these occurrences and show why one very popular way of thinking about them is not fully adequate. The need to address the failings of that proposal will then serve as the jumping-off point for the next chapter.

In §1, I review some basic facts about anaphora. I claim that anaphora is conventionally used to informationally organize a conversation to ensure its participants are suitably coordinated on the entities it concerns. By this, I mean that it is a way of clustering together the occurrences of certain expressions into discrete groups and that competent interpreters are required to understand that each group corresponds to a given subject of conversation under discussion. The notion of a discourse referent is then introduced to explain how this is accomplished. At a minimum, discourse referents serve as labels for clusters of manditorilly co-designating expressions of this kind. More substantive interpretations can then serve as the basis for explaining the basic mechanics of how this clustering is achieved in the first place.

In §2, I discuss a prominent and relatively conservative way of developing an interpretation of that kind. In particular, certain expressions appear to have singular use-conditions, in the sense that speakers are normally required to have specific individuals in mind when they deploy them in conversation. And if this is right, these individuals could plausibly be identified with the intended subjects of an utterance. It follows that discourse referents could then be identified with individual concepts determined by the communicative intentions of the speakers that introduce them.

In §3, I argue that this attractive way of approaching the issue, unfortunately, fails

to yield a theory that works in full generality. More specifically, it appears incapable of dealing with a family of cases that I call anaphora under operators. These cases all still exhibit the essential discourse-structuring hallmarks of anaphora, even though the expression occurrences introducing the novel anaphoric possibilities are most naturally interpreted as taking scope under some kind of structurally superior operator. It follows that in each of these cases, there is not an individual that the speaker could have plausibly had in mind as the corresponding subject of their utterance, and so traditional individual concepts will be an insufficient tool for modeling the discourse referent at work.

3.1 Anaphora and the Informational Organization of Discourse

A persistent theme in the philosophy of language has been to reflect on the interesting properties enjoyed by certain special patterns of co-reference. More specifically, some strike us as somehow logically or semantically guaranteed to hold, which makes them stand out from others that strike us as holding in a less secure manner. Frege (1892) famously illustrated the point by contrasting two different identity statements:

(16) a. `Hesperus is Phosphorus.`

b. `Hesperus is Hesperus.`

There is a pretty obvious epistemic distinction between (16a) on the one hand and (16b) on the other, even though both seem to be representationally equivalent (in the sense that both ascribe the same relation to the same pair of entities). Since (16a) is used to express a reasonably substantial astronomical discovery, being in a warranted position to assert it seems to require at least *some* knowledge of the arrangement of the heavenly bodies. But warranted assertion of (16b) requires substantially less – its truth seems to be *logically* guaranteed so long as `Hesperus` succeeds in referring to anything at all.

Following some recent terminology, we may call the kind of ordinary co-reference exhib-

ited by (16a) *de facto*, and the more cognitively interesting form of co-reference exhibited by (16b) *de jure*. While there is still a good deal of controversy about its precise formulation, the special epistemic feature of *de jure* co-reference that sets it apart from *de facto* co-reference can at least be roughly formulated as follows: it is a condition on the linguistic competence of an interpreter that she be in a position to recognize and appreciate that the occurrences of *de jure* co-referential expressions are in fact co-referential. Failure to recognize co-reference in these cases is a manifestation of a more fundamental defect in their knowledge of linguistic conventions and practices, not a defect in their knowledge about the world more broadly.[1] Our starting place is the observation that natural language appears to have a special-purpose device for conventionally indicating something like *de jure* co-reference. In particular, *anaphoric pronouns* appear to be *de jure* co-referential with their antecedents when those antecedents are singular terms:[2]

(17) a. `Sam Clemens thought that Mark Twain wrote great short stories.`

b. `Sam Clemens`[1] `thought that he`$_1$ `wrote great short stories.`

Since, as a matter of fact, `Sam Clemens` and `Mark Twain` are used to refer to the same person, these two terms are co-referential in their appearances in (17a). This fact, however, does not need to be appreciated in order to understand what (17a) says. Someone could, for instance, be familiar with the author in a few different ways while also failing to appreciate the fact that those are ways of thinking about one and the same person. If that were the case, there would be no reason to count them as somehow incompetent with the uses of the names `Sam Clemens` and `Mark Twain`, and so this quasi-factual misunderstanding would still be no impediment to them understanding the fact that (17a) reports. Contrast that with what is happening in (17b). There we likewise seem to have two elements which may be said

[1]This statement is intended to be more or less theory-neutral, and so the underlying notion of knowledge of language at issue should be understood in a broad enough way to be uncontroversial.

[2]The observation that the semantic significance of co-indexation in the case anaphoric pronouns is to signal something like *de jure* co-reference is often made in passing, but it is the central theme of Fiengo and May (1994, 2006).

to co-refer – `Sam Clemens` and the anaphorically linked pronoun `he`. But this form of co-reference arguably does need to be appreciated to count as understanding what (17b) says. In other words, successfully interpreting (17b) (but not (17a)) in part *requires* recognizing that the believer and the subject of the reported belief are one and the same individual.

For our purposes, it is important to first realize that these kinds of anaphoric dependencies can manifest themselves in linguistic units that are larger than single utterances:

(18) `Mark Twain`[1] `was proud of his`$_1$ `work. He`$_1$ `thought that he`$_1$ `wrote great short stories.`

(19) **A**: `Lucy Lawless`[1] `is an extremely popular actress.`

B: `What characters has she`$_1$ `played?`

In (18), we see them stretch across a sentence boundary, and in (19), we see them crossing *both* a sentence *and* speaker boundary. Even in these extended situations, the type of co-reference is apparently *de jure* and not *de facto*. Thus, an interpreter should be counted as failing to understand (18) if he does not recognize that both sentences are about the same man. Likewise, part of understanding the brief dialogue in (19) is recognizing that the woman whom **B**'s question is about is the same as the one that **A**'s assertion was about. Anaphoric dependencies like these thus seem to play an essential role in the organization of discourse. They are a way of ensuring that a large body of distinct utterances is suitably structured so that they are guaranteed to be understood as being about the same subjects.

In fact, this discourse-structuring role seems to be the *real* conventional significance of anaphoric pronouns. This point is best appreciated by noticing that something like this same general phenomenon also manifests with what are apparently *non-referential* terms. Most mainline theorists will agree that one of the essential semantic functions of the indefinite article is to signal existential quantification of some kind. But despite the overwhelming and long-standing agreement that statements involving indefinite noun phrases have existential truth-conditions, some uses of indefinites also seem to lack one of the major traditional hall-

marks of true quantification. Most famously, it is sometimes extremely difficult to apply the traditional notion of the *scope* of a quantifier to an indefinite noun phrase. On standard analyses, the scope of a quantifier expression can be syntactically defined as whatever that expression structurally commands at the semantically relevant level of syntactic description. They are then predicted to only have a semantic effect within that syntactically delimited domain. But certain uses of indefinites appear to project semantic dependencies far beyond any plausible syntactically delimited scope boundary. That point is most strikingly illustrated by occurrences of indefinites that participate in an anaphoric dependencies in longer stretches of discourse:

(20) a. `Bill owns a car`^{1}`. It`$_1$ `is black.`

b. $\exists x(\text{CAR}(x) \wedge \text{OWNS}(\mathbf{b}, x) \wedge \text{BLACK}(x))$

(21) **A:** `A burglar`^{1} `broke into our office.`

B: `Oh no! What did he`$_1$ `steal?`

The sequence of utterances in (20a) is perfectly coherent and plausibly results in the truth-conditions indicated in (20b). But the formal scope configuration in (20b) is completely implausible as a faithful syntactic representation of the structure of the sequence in (20a) since the existential quantifier would have to locally take scope only in the sentence in which the indefinite actually occurs. The brief dialogue in (21) illustrates that the problem is even somewhat worse than that: here, **B**'s question involves an occurrence of an anaphoric pronoun that is co-indexed with an occurrence of an indefinite in **A**'s earlier assertion. Assigning an adequately wide enough scope to the indefinite in this case would involve saying that it took effect over the utterances of two different speakers making what appear to be two different kinds of speech act. And in this case, it is unclear how to even begin to formally represent that kind of scope configuration, let alone in a way that is somewhat faithful to natural grammar. So it seems better to think that indefinites are like proper names in the sense that they are *scopeless*, though still somehow also managing only to

contribute an existential quantifier to truth-conditions.

The general puzzle raised by the apparent scopelessness of indefinites despite their existential truth-conditional contribution has generated an enormous body of literature trying to work out the exact interpretive mechanisms which allow them to have both of these features in a broadly compositional framework. But we can safely ignore the exact details of those proposals at the moment and instead just appreciate that this puzzle only gets started because we are willing to accept that the patterns of co-indexation in examples like (20) and (21) are legitimate. In the present context, the grounds for that claim cannot directly have anything to do with *de jure* co-reference since the other ingredient of the puzzle requires that indefinites be non-referential expressions (and so they cannot be co-referential with their linked pronouns). Instead, it seems like we need to appeal to something like an informal notion of a discourse's subject matter. Both of the discourses in (20) and (21) are intuitively regarded as being about some object – in (20), the utterances are all about the same car, and in (21), they are all about the same burglar. This intuitive sameness of subject matter is then what explains our willingness to say that the discourses in (20) and (21) are structurally parallel to those in (18) and (19). All of them are informationally structured in broadly the same way so that they each must be understood as being about a single subject. The apparent *de jure* nature of the co-reference realized in (18) and (19) is then explained by this general structural fact *together with* the fact that the terms related by that structure in (18) and (19) happen to be referential.

Early work in the study of anaphoric connections like this sometimes goes under the heading of *binding theory*. One of the main goals of traditional binding theory was to provide a reasonably concise (and ideally also purely linguistic) characterization of the distribution of anaphoric elements in a given language since one of the more striking structural facts about human languages is that the admissible patterns of antecedent/anaphor relations are generally highly constrained:

(22) a. $\texttt{Sabrina}^{1}$ `talks to her`$_{1}$ `cat.`

b. *She[1] talks to Sabrina's$_1$ cat.

(23) Bill doesn't own a car[1]. *It$_1$ is black.

The sentence in (22a) may be understood as being about Sabrina, while the sentence in (22b) seems like it must be understood as being about two potentially different women – Sabrina and someone else. Thus, the indicated patterns of co-indexation are licensed in the case of (22a) but not (22b). Similarly, it is hard to understand the first sentence of (23) as being about anything other than Bill. So the indicated pattern of co-indexation is ruled out since it suggests that this sentence is also about a particular car that Bill does not have. Binding theory tried to state very general principles that allowed you to derive the kinds of constraints at work in these examples.

An influential early paper in this tradition stated some of these core questions in an extremely vivid way that is particularly well suited for our present task. Karttunen (1976) begins by asking us to consider a reasonably concrete language processing task and reflect on some of the capacities that an idealized machine for reading English text and storing the information it extracts would have. Such a device would minimally be expected to keep a running list of the objects mentioned in the text that it is currently reading and answer any questions about those objects that are settled by it. In order to do this job well, every time the machine encounters a designating expression, it needs to be able to determine whether the object it designates is *potentially novel* (thus requiring that it open a new database for storing information about it) or had already been *previously mentioned* (in which case an older database would need to be updated with new information). In this context, it is pretty clear that the semantic significance of anaphora is to signal some kind of further elaboration on a previously mentioned entity. According to Karttunen, an occurrence of an expression is said to introduce a new *discourse referent* just in case it introduces a novel anaphoric possibility. So when the interpreting machine encounters an occurrence of an expression that introduces a new discourse referent, it will be expected to construct a new database for storing information about the object it designates. On this way of looking at things, the

central task of binding theory is determining which occurrences of expressions introduce new discourse referents and which need to reuse old discourse referents.

It is important to realize that Karttunen's use of discourse referents is completely ontologically neutral. In a given discourse, co-indexation is obviously an equivalence relation, and the discourse referents present in a discourse correspond to the equivalence classes of that relation in a one-to-one fashion. It is tempting to see this as the basis of a purported explanation of the co-indexation relation. Two expressions are co-indexed *in virtue* of the fact that they correspond to the same discourse referent.[3] This way of understanding the situation would surely take on a significant ontological commitment since discourse referents are being used in a non-trivial way to ground certain facts concerning an important theoretical relation. But while this general picture is very attractive, it is not obviously forced on us. For example, we may take facts concerning the co-indexation relation to either be primitive or grounded in some other way and view discourse referents as a harmless abstraction of that relation. And in that case, they would have an ontological commitment to them in only the thinnest of senses. Karttunen's use of the notion is at best undecided between these two positions, if not explicitly of the latter variety. In particular, the only specific criteria for the individuation of discourse referents that we have been given cuts through considerations of co-indexation.

The theoretical notion of a discourse referent is clearly related to the intuitive notion of a discourse's subject matter that I have been appealing to so far. But it would be a mistake to assume that the relation between them is one of straightforward refinement. Even though discourse referents have been provided only relatively thin individuation conditions, those

[3]You could worry about whether or not this general strategy is actually appropriate. While co-indexation is clearly an equivalence relation, you might think that there are more fundamental and important relations that are not, which should be our real target of analysis. The recent debate about whether or not *de jure* co-reference is an equivalence relation is a case in point. For instance, Pinillos (2011) maintains that the *de jure* co-reference relation is non-transitive, and hence any strategy for explaining that relation that appeals to the literal identity of some shared third-object is doomed to fail. While I am very sympathetic to these concerns, they are somewhat orthogonal to the task at hand, and so I will set them aside.

are enough to start deviating from intuition in certain key ways:

(24) `A dog`[1] `broke into my yard last night. It`$_1$ `ruined my herb garden!`

(25) `You have two messages: a colleague`[1] `wants you to read her`$_1$ `paper, and a friend`[2] `wants you to meet her`$_2$ `at the pub.`

Common sense says that a speaker is licensed to utter the sequence in (24) on a relatively thin informational basis: a ruined herb garden, together with some background knowledge about the irritating habits of the neighborhood pets, is surely enough. Against this informational backdrop, we are tempted to say that the intuitive subject of (24) is *whatever* happened to leave the evidence that licenses the initial assertion.[4] But note that this informational basis is not enough to ensure that there is a unique entity who committed the crime. And that is perfectly fine – we still have the intuition that an utterer of (24) speaks truly even when, perhaps, it was a pair of dogs that made the intrusion. But there is only a single discourse referent at work here and no obvious fact of the matter about which of the two dogs that discourse referent corresponds to. So it appears that we have a one-many mismatch between the theoretically individuated discourse referent and the intuitive candidate discourse subjects. The sequence in (25) illustrates the potential for a numerical mismatch in the other direction. Suppose an editor of a journal leaves her office for lunch and so has her assistant take her calls. Upon her return, he might utter (25) on the basis of two different calls he fielded. This sequence involves two distinct discourse referents. But it is not very difficult to imagine a situation where the colleague and friend in question are, in fact, the same person, which the assistant has simply failed to appreciate for whatever understandable reason. In that case, we are inclined to say that the sequence in (25) is really about that person, whoever she may be. But then we seem to have a many-one mismatch between the theoretically individuated discourse referents and the intuitive candidate discourse subject.

[4]I take it that this is the main intuition driving the so-called *de fonte* account of discourse subjects put forward by Zimmerman (1998) and further developed by Dekker (2000).

I take the above considerations to be more or less decisive against the view that discourse referents can be used to *directly* characterize the theoretical role of discourse subjects. But a modestly more substantive view of discourse referents will be able to help us indirectly characterize that role in a satisfactory manner. We can start by noting that the relevant instance of the sequence (25) discussed above is obviously just another case of the oft-discussed phenomenon of double-vision. The issue is that an individual may (to use a familiar manner of speaking) be familiar with one and the same object under several different guises in a way that does not render them capable of transparently making true identity judgments. In this case, the assistant was assumed to think of the same object in roughly two different ways: one *qua* his superior's colleague, and one *qua* her friend. The discourse referent-count, in this case, thus seems to agree with the intuitive guise-count. That suggests that discourse referents could stand to a discourse's subject matter in something like the relation that individual concepts are thought to stand to individuals. Of course, as (24) indicates, things cannot be quite that simple – at a minimum, we need to potentially allow for the possibility of one discourse referent standing in the relation to many distinct subjects. But I think the analogy is a helpful one, and so in what follows, I will adopt the perspective that discourse subjects are something like the (plural) *designata* of discourse referents as a working hypothesis.

This way of looking at things seems to *prima facie* rule out an entirely instrumentalist attitude towards discourse referents. At a minimum, we are owed an account of the designation relation, and it is not clear how to adequately do that without accumulating at least some specific commitments about their nature that go beyond their mere individuation in terms of co-indexation. What follows should thus be understood as experimentally adopting a more inflationary attitude towards them than is sometimes assumed. But it is not clear exactly how inflationary we really need to be. In particular, we may be able to model discourse referents using theoretical machinery that we are already antecedently committed to and already has a well-established track record in semantic and pragmatic theorizing. Since

a conservative proposal of this kind is clearly preferable to one that requires adopting some new theoretical primitives, it will be useful to have a view of this kind on the table as a kind of default position. So I will now attempt to describe a reasonably popular position with approximately that character.

3.2 Discourse Referents as Individual Concepts

One of the central challenges that needs to be overcome in developing an account of this kind is appropriately handling the above-mentioned peculiar features of indefinite noun phrases. That is because they may freely participate in long-distance anaphoric relations of the kind we are interested in, and a working understanding of their core interpretive behavior seems like a precondition on being able to characterize the discourse referents they introduce adequately. One of the simplest ways to explain why indefinites semantically pattern in certain respects with referential rather than quantificational expressions is to bluntly assume that they, in fact, *are* referential expressions in a certain sense. This kind of suggestion has a long pedigree, and while it is now well-known that a simple ambiguity between specific (referential) and non-specific (quantificational) uses of indefinites cannot predict the full range of semantically anomalous behavior that indefinite noun phrases appear to participate in, it is still a *prima facie* plausible account of the particular use of indefinites we are currently considering (i.e., those in which the expression has apparently unlimited scope over the discourse).[5]

[5]The apparently referential character of certain uses of indefinites was pointed out in the ending sections of Strawson (1950). Fodor and Sag (1982) developed a more detailed linguistic proposal which posited a specific/non-specific ambiguity for indefinites that is not unlike the referential/attributive ambiguity for definites noted by Donnellan (1966). The general problem with this kind of view is that some uses of indefinites can take special scope in certain positions *within* a sentence that seem to also resist a straightforward syntactic treatment. These uses cannot be explained with a simple dichotomy between specific and non-specific uses since here they make an obviously quantificational contribution to truth conditions while still assuming a scope position that is unavailable to most other kinds of quantifier expression. The point mirrors the observation in Kripke (1977) that Donnellan's referential/attributive distinction cannot account for the same range of readings that Russell's notion of the scope of a definite description can. See e.g. Reinhart (1997) and Winter (1997) for discussion.

It has been emphasized by a number of authors that a seeming felicity condition on this use of indefinites is that the speaker have certain *referential intentions*.[6] Here is an example which clearly illustrates this point:

(26) **A**: `A student`[1] `came by our office looking for you.`

B: `Did she`$_1$ `have black hair?`

Like (21), the exchange in (26) features a use of an indefinite that clearly exhibits the relevant anomalous interpretive properties in a rather strong form. Suppose the facts of the situation are as follows. In fact, *two* of **B**'s students both stopped by **A** and **B**'s shared office at different times looking for her (though **A** was perhaps only aware of one). One of them has black hair, and the other does not. **A**'s utterance is clearly true in this situation – that is certainly not in doubt. But just as strong is the intuition that B's question is *perfectly well formed* and must have a definite answer. What that answer is, of course, depends on some facts about the situation that we have not provided. More specifically, it depends on which office visitor **A** had in mind when he made his utterance. He is, of course, free to have either, and so both *yes* and *no* may be correct answers depending on what his initial communicative intentions were.

Two things are worth emphasizing about this case. The first is that the condition that **A** has an intention to refer to some particular student when he made his utterance is not optional. In other words, he is not free to reject **B**'s question as ill-formed because he did not have any one student in mind. So even if you think that existential quantification exhausts the *truth-conditional* contribution that an indefinite expression makes, that alone is not sufficient to predict the full range of interpretive facts concerning their use. The second is that which student **A** has in mind may have some kind of truth-conditional relevance downstream. In particular, **B**'s use of the pronoun `she` seems stipulated to have as referent whomever that individual is. Hence, the question that she is asking seems in part to depend

[6]See e.g. Stalnaker (1998), Dekker (2002), and Cumming (2015).

on what individual **A** happens to have in mind since the truth of a simple *yes* or *no* answer in response to that question depends on whomever that individual is. If you had the one with auburn hair in mind, but due to the poor lighting conditions in your office misperceived it as black, then the answer *yes* seems strictly false, despite it being true that a student with black hair did come by looking for **B**.

The phenomenon of so-called *pronominal contradiction* also serves to highlight the specific use conditions on these occurrences of indefinites in an extremely strong form:

(27) **A**: `John`^{1} `just threw out an old broken typewriter`^{2}`.`

B: `It`$_2$ `wasn't a broken typewriter -- it`$_2$ `was a brand new computer. And he`$_1$ `did not throw it`$_2$ `out -- he`$_1$ `donated it`$_2$ `to charity.`

Notice that **B** is allowed to use the pronoun `it` in his utterance in a way that is strongly coordinated with **A**'s use of an indefinite, but in an extremely odd sense. The dialogue above *must* be understood as concerning some definite object to which **A** intended to refer. But **B**'s disagreement with almost every aspect of **A**'s utterance indicates that it cannot be identified on any putatively linguistic grounds since **B** is willing to reject essentially all of the explicit descriptive information that **A** has assigned to the object in question. Instead, her use of the pronoun seems like it is best understood as being stipulated to designate some definite object that **A** had in mind in making his utterance, and which she can identify on independent grounds.

Stalnaker (1998) takes the above observations about the specific use conditions of indefinites as the basis for what I take to be a reasonably conservative theory of discourse referents and discourse subjects. In most of his work on pragmatics and the semantics/pragmatics interface, Stalnaker appeals to a construct that he calls the *common ground*. Informally, the common ground is the set of propositions that the conversational participants mutually accept for the purposes of conversation.[7] This is then coupled with Stalnaker's characteris-

[7]In order to avoid certain problem cases, acceptance for the purpose of conversation is taken to be its

tically austere conception of a proposition. A sentence's propositional content is exhausted by its truth-conditions, which may be modeled in the typical way as a set of possible worlds. The result is an extremely austere view of context, as well. All of the information carried by the common ground can be represented by the *context set*, which is something like the conjunction of every proposition in the common ground. Intuitively speaking, the context set is really just the strongest proposition mutually accepted by the participants in the discourse. Stalnaker's broader project in this area in part involves trying to show that this proposition (conceived of as a mere set of epistemic possibilities) is a rich enough body of information to bear most if not all of the explanatory burden in pragmatics. In other words, he is interested in showing that the context set is the only notion of context needed to explain any theoretically tractable issue regarding situated language usage.

The fact that indefinites appear to introduce discourse referents has sometimes been used to argue against this aspect of the Stalnakerian program. One familiar version of that argument starts by considering the following pair:

(28) a. `A student`^{1} `was looking for Albert. She`$_1$`...`

b. `One or more students`^{1} `were looking for Albert. *She`$_1$`...`

There is clearly an important interpretive difference between the sequences in (28). The singular pronoun in (28a) is naturally and automatically understood as having a certain student as its denotation, but for whatever reason, this interpretation is blocked in (28b). Using some of the theoretical vocabulary introduced in the previous section, we may say that the indefinite expression `a student` introduces a discourse referent that has that student as its designatum, while the expression `one or more students` does not. The Stalnakerian view of content and context seems to be unable to account for this difference. We are to imagine two scenarios in which the sentences in (28) are each respectively uttered against

own *sui generis* propositional attitude so that an agent might count as accepting a proposition that she does not believe or believing a proposition that she does not accept.

the backdrop of the same background context. Assuming that an indefinite article only semantically contributes an existential quantifier, then truth-conditions of the utterances in both of these scenarios are the same, and so according to Stalnaker, the same proposition will have been asserted. So if we follow Stalnaker (1975) in taking the main interpretive effect that an assertion has on context to be the addition of the asserted proposition to the common ground, it seems like the context that results from the utterances in these two scenarios is also going to be the same. But this obviously is not right: the context will afford speakers a certain anaphoric possibility when they utter (28a), but not when they utter (28b).

Stalnaker's response to this objection is that it incorrectly assumes that the addition of the asserted proposition is the *only* relevant change to the common ground brought about by the assertion. But as noted above, the use of an indefinite expression like `a student` is often presumed to carry with it referential intentions on the part of the speaker. This aspect is not obviously shared by other quantifier expressions like `one or more students`. Hence, there is a certain fact of the matter operative in any discourse in which an indefinite of the relevant type has been used that does not hold when a truth-conditionally equivalent quantifier expression has been used instead. Specifically, there is a fact of the matter about which intended object licensed the use of that indefinite expression. Assuming the interpreter is generally aware of this difference, it follows that the proposition which asserts the existence of such a fact would be accommodated into the common ground after (28a) was asserted, but not after (28b) was asserted. So, the resulting context is not exactly the same. The important difference between them is that every world in the context set after an assertion of (28a) is one for which a certain metalinguistically determined individual concept is defined, which we may approximate as *the individual to whom the speaker intended to refer*. This concept is *not* guaranteed to be defined at every world in the common ground after an assertion of (28b). Thus, a speaker may safely use that concept as a mechanism for fixing the denotation of the pronoun occurrence in (28a), but not (28b).

While he might not put it this way himself, we can see in Stalnaker's response the outline

of a theory of discourse referents and the subjects they designate. We begin by noting that certain occurrences of expressions have *singular use-conditions*, meaning that the speaker needs to loosely have a certain object in mind in order to be warranted in deploying them. This is more or less obvious for most occurrences of proper names, where those preconditions on their appropriate use are roughly aligned with their truth-conditional contribution in the sense that the object the speaker has in mind is ordinarily required to be the same as the name's direct semantic contribution.[8] But the use-conditional and truth-conditional aspects of an occurrence of an expression can also sometimes be *misaligned*, as witnessed most obviously by the kinds of positive occurrences of indefinites we have been concerned with. Here, despite having singular use conditions, they only semantically contribute an existential quantifier.[9]

[8]The only reason that I am hedging here is because of certain puzzle cases due to Kripke (1977), which suggest that sometimes the object a speaker has in mind when she uses a proper name may be distinct from that name's real semantic contribution. But the circumstances where this may arguably happen are rather special, and so we may continue to safely endorse the alignment at least as a plausible default without having to come down one way or another on those examples.

[9]The explanation for this kind of divergence presumably has to do with the biases that these two different vantages put on either production or interpretation. The use-theoretic perspective is more-or-less entirely speaker-oriented. It is concerned with the kind of evidence that a speaker needs to be in possession of in order for her assertions to be licensed. The truth-theoretic vantage, on the other hand, is arguably more audience-oriented. It is principally concerned with characterizing some workable standard of correctness that can be applied by an *interpreter* in evaluating the appropriateness of any encountered utterance. The difference in emphasis between these two perspectives is sometimes overlooked because they are often seen to be interchangeable in the context of other theoretical principles. The rough idea is that use-conditions can be freely converted to truth-conditions, and truth-conditions give rise to (certain) use-conditions via general bridging principles like the truth-norm on assertion. The match is obviously not one-to-one, but there are enough bridging principles of this kind that are standardly adopted to make this difference in principle hard to detect and adjudicate.

But indefinites appear to solve a communicative task that makes this imperfect match a little easier to see. Singular thought is an extremely widespread phenomenon, and so it seems perfectly possible for agents to be in possession of certain pieces of singular information for which they lack a ready-at-hand mechanism to articulate. Suppose, for instance, that an agent is in possession of some singular information about an object they are (or were) perceptually acquainted with and, for whatever reason, want to share it with their conversational partner. Moreover, assume that for whatever reason, the speaker is not in a position to assume that her conversational partner is acquainted with the relevant object in some way or another. Standard theories regarding the semantics of singular reference actually look like they put this speaker in a pretty awkward position. A simple case of singular demonstrative reference is out of the question since the object in question is out of view. And if there is no name or uniquely-identifying descriptive condition ready at hand, the speaker seems almost completely out of luck for even communicating a related general thought.

Surely this kind of situation is common enough for it to be reasonable for a language community to develop

According to this perspective, the objects that license the use of these expressions are the subjects that a discourse concerns. Each subject is therefore contingently associated with an occurrence of a particular expression. That association enables an agent to think about a discourse subject *qua* the licenser of the expression occurrence that it is contingently associated with. Moreover, these guises can be easily modeled as individual concepts of the familiar sort – that is, as a partial function yielding in each world for which it is defined the licenser of a given occurrence of an expression with singular use-conditions. Importantly, an interpreter will often not be in a position to independently identify the licenser of an expression with singular use-conditions since their epistemic state will often be compatible with several different candidate witnesses. It follows that these concepts are usually non-rigid, and so their designatum will normally vary from world to world. This feature makes these individual concepts well suited to play the theoretical role of discourse referents since two non-rigid individual concepts may agree on their value in the actual world (and hence designate the same subject) while still being strictly distinct. Discourse referents so construed can then be deployed to interpret the pronouns co-indexed with the expressions that introduced them.[10]

---

some conventional means of addressing it. The particular use of indefinites that we have been considering seems to be just such a device. The use-conditions for these occurrences of indefinites appear to be keyed to the fact that the speaker intends to convey some piece of singular information. But without uniquely identifying descriptive conditions ready at hand, the speaker must instead opt to provide only partially identifying conditions. In reaction to such a move, the audience is then put in the following position: they would know that the speaker intended to convey a singular thought within some grammatically demarcated range but would be unable to select among the propositions in that range. Thus, the most informative proposition they could help themselves to would be the *disjunction* of that range, and so the expression would naturally be assigned existential truth-conditions from their perspective.

[10]There are two familiar options for how this might go: the relevant individual concept could give the meaning of the pronoun, or it could merely be used to fix its designation. The latter option seems to be the more natural one. Despite in a certain sense being non-rigid, they appear to semantically behave like so-called pure indexicals. As Kaplan (1989a) pointed out, pure indexicals like the first person pronoun clearly have certain descriptive and context-invariant meanings. In the case of the first-person pronoun, in each context it functions to pick out the speaker of that context. But that descriptive aspect of its meaning appears to be semantically inert since no semantic operators have access to it – their designation cannot be shifted by embedding the expression in a larger intensional construction. This aspect of their semantic behavior is explained by saying their descriptive content only serves to fix the designation of its occurrences and does not literally give the meaning of those occurrences. The occurrences of anaphoric pronouns of the kind we have been discussing here also arguably resist embedding of this kind (the only counterexamples

Stalnaker's explanation of the contrast in (28) nicely illustrates how this theory works when it is applied to one of the simplest anaphoric configurations that it is accountable for. It also seems to deliver the correct results for some of the more complicated examples that we have discussed. For instance, let us again consider (25) in this light:

(25) `You have two messages: a colleague`^{1} `wants you to read her`$_1$ `paper, and a friend`^{2} `wants you to met her`$_2$ `at the pub.`

To remind the reader, the circumstance that interested us is one where (25) is uttered by a speaker who was unaware that the colleague and friend in question are, in fact, one and the same individual. In this case, we are naturally inclined to say the speaker, in some sense, had the same individual in mind when they used the indefinites `a colleague` and `a friend`, though perhaps he had her in mind in different ways. But however we want to gloss the situation, the resulting discourse referents these indefinites introduce are inarguably distinct. The individual concept that maps a world to the object the speaker had in mind in using `a colleague` in that world is clearly not the same as the one which maps a world to the object the speaker had in mind in using `a friend`, since it is not a necessary fact that the two objects picked out in this way are identical. And the result is that an audience member who uses these concepts to interpret the respective co-indexed pronouns will be in the right epistemic state. By this, I mean that they will be guaranteed to transparently understand the tight interpretive connection that holds between the two elements of each co-indexed pair but will not be guaranteed to transparently understand the identity that holds between subjects designated by the two pronouns. Similarly, the account seems to easily make sense of the guaranteed form of cross-speaker coordination witnessed by (26), since the denotation of **B**'s use of the pronoun `she` is stipulated to be whatever **A** had in mind. And it can even

---

that I am aware of involve embedding under propositional attitudes, but other pure-indexicals also seem to be ill-behaved in those environments as well). This will sometimes leave us in an awkward position, as intuitive judgments about the information conveyed by an utterance are often better captured by thinking that discourse referents give the meanings of anaphoric pronouns instead. But at this point, Stalnaker may help himself to some of the other aspects of his 2-dimensional pragmatic framework. In particular, diagonalization may be used as a pragmatic repair strategy to account for counterexamples of this kind.

seem to account for cases of pronominal contradiction like (27), since **A** and **B** may disagree about which properties the object **A** has in mind possesses.[11]

Of course, the above sketch also leaves many questions open. For instance, it is unclear how to extend it to handle cases like (24):

(24) `A dog`[1] `broke into my yard last night. It`$_1$ `ruined my herb garden!`

Here, the issue was that an utterance of (24) could be warranted even in information states where the speaker cannot strictly rule out the possibility of more than one witness for the existential claim. Moreover, supposing that there are, in fact, multiple witnesses to the claim, that situation can be one in which the speaker does not have any obviously privileged epistemic relation to any specific one of them. All of the relevant informational channels she has exploited in coming to believe the proposition that her utterance expresses may apply equally well to each of the witnesses in question. It follows that there is not any obvious answer to the question of *which* of these objects she had in mind when speaking, and so no obvious answer to the question of what the subject of the discourse is on the present proposal. In response, we could make one of three moves. We could simply accept the result, commit ourselves to the claim that her utterance is in a certain sense defective, and deny that the pronoun receives a semantic value in this kind of situation. While some aspect of this proposal is surely right (we do not want to say that *nothing* has gone wrong here), it is also extremely difficult to shake the feeling that she has still spoken truly. Thus it seems preferable to instead either say that in this case, she has all of the relevant candidate

[11]In fact, the ability to handle pronominal contradiction easily seems to be a decisive reason in favor of Stalnaker's version of the conservative strategy. Another popular approach due in essence to Evans (1977) and popularized by Neale (1990) similarly takes the discourse referent/discourse subject dichotomy to be something like the individual concept/individual dichotomy but understands these concepts as being given by some overt linguistic material rather than a speaker's communicative intentions (this is not *quite* right as both Evans and Neale adopt Russell's theory of descriptions, and so it is a little distorting to state their view in terms of individual concepts. Putting it this way more accurately describes the view recently defended in Elbourne (2013), which is extremely structurally similar to Neale's theory but instead adopts a Fregean semantics for definite descriptions.) Pronominal contradiction appears to be a straightforward counterexample to this kind of view, as there is not obviously any suitable overt linguistic material that can be used to get the right results.

witnesses in mind (in which case we would need to generalize our proposal to allow for plural discourse subjects), or that despite appearances to the contrary, there is a fact of the matter about which one she had in mind (perhaps by appealing to something like the device of arbitrary reference presented in Breckenridge and Magidor (2012)).

This problem seems to point to a more general issue. The theory has adopted some traditional vocabulary from the theory of singular thought and freely appealed to the relation that holds between agents and the individuals that they purportedly have in mind. But it did not provide any principled characterization of that relation, and being in a position to adequately choose between some of the above options clearly presupposes that we have something like a working characterization ready at hand. So completely fleshing out the theory seems like it would inevitably require a detour through the literature on singular thought. Without a specific account of this core representational relation, the present proposal can only be regarded as a bare skeleton. But it does seem like a promising and attractive one. In particular, it accords with intuition on many key cases and only uses theoretical primitives that we take ourselves to be in need of anyway. We have also yet to see any serious obstacles to its development. At worst, it will possibly deliver a counter-intuitive verdict in certain highly specialized fringe cases. It thus seems to meet all of the criteria that we would want a good theoretical default to meet. Unfortunately, as you will soon see, I believe that there is good reason to deviate from this theoretical default.

3.3 Anaphora Under Operators

The most compelling counterexamples to the view of discourse entities presented in the previous section all fall under the general heading of what I will call *anaphora under operators*. These are situations where the existential interpretive contribution made by an indefinite is most naturally regarded as taking effect under the scope of some structurally superior operator, but we are still willing to endorse a strong interpretive connection between that

indefinite and a later pronoun which appears to be well outside the scope of that operator:

(29) A wolf[1] could walk in at any minute. It$_1$ would eat you.

(30) Hob thinks a witch[1] has blighted Bob's mare, and Nob wonders whether she$_1$ killed Cob's sow.

(31) Harvey courts a girl[1] at every convention. She$_1$ always comes to the banquet with him.

As Roberts (1989) points out, the most natural interpretation of the first sentence of (29) assigns the indefinite narrow scope with respect to the possibility modal. It follows that this occurrence of an indefinite does not plausibly have singular use-conditions, if we construe that as meaning that the speaker needs to have a particular wolf in mind when making her utterance. If she *did* have some wolf in mind, it seems she would instead be trying to communicate a stronger proposition where the indefinite is assigned a wider scope. The problem is that this occurrence of an indefinite still appears to introduce a discourse referent since it licenses an occurrence of a co-indexed pronoun later in the utterance, which is clearly outside any plausible grammatical scope boundary that the indefinite might have. But without singular use-conditions, the Stalnakerian theory seems powerless to characterize that discourse referent.

Some of the examples that Geach (1967) used to illustrate the phenomenon of *intentional identity* raise similar problems. Geach found (30) interesting because it can be used to truly express a kind of attentional coordination between Hob and Nob, which we might naively gloss by saying that they are both thinking about the same witch. But this remains true even in situations where there are, in fact, no witches to be thought about. The challenge is to give a positive account of what their coordination consists of if it is *not* joint attention on a particular (existing) entity. For our purposes, we just note that the circumstances that interested Geach are ones where the indefinite in (30) must take narrow scope with respect to the attitude operator. If there are no witches, there cannot be a particular witch that

Hob is thinking about, and so his belief must be understood as being *de dicto* and not *de re*. But this indefinite is co-indexed with a latter occurrence of a pronoun that is seemingly outside the scope of that operator (since it is under the scope of an ostensibly different attitude operator). It thus seems like we must understand this indefinite as introducing a discourse referent. But the Stalnakerian account of that discourse referent again seems *prima facie* ruled out. A speaker may know perfectly well that witches do not exist and yet still truthfully and felicitously utter the entire sequence in (30). So we cannot demand that the speaker have a particular witch in mind when making that utterance, and so likewise cannot define a discourse referent in terms of those singular intentions.

Dependent anaphora is yet another example of what I take to be the same general phenomenon. The sequence in (31) is due to Karttunen (1976). The first sentence in this sequence contains two different quantifier phrases and is hence ambiguous between two readings corresponding to their two possible scope orders. The relevant reading for us is the one that grants the universal quantifier scope over the indefinite. This reading is compatible with Harvey courting different girls at each convention, so long as he courts at least one. Somewhat surprisingly, even this occurrence of an indefinite will introduce a discourse referent. The occurrence of the co-indexed pronoun in the second sentence of this sequence is still perfectly compatible with that reading. The kind of anaphora exhibited in (31), therefore, also causes problems for the Stalnakerian theory. We cannot explain the discourse entities at work here by appealing to singular communicative intentions of some kind since the speaker, in fact, does not appear to need to have a particular individual in mind when uttering of (31) if she merely intends to communicate the reading which gives the indefinite narrow scope.

We should resist the temptation to treat these cases as somehow being special exemptions to the general rule that occurrences of indefinites that introduce discourse referent have singular use-conditions. First, note that the significance of co-indexation in these cases still seems to broadly serve the same role in securing some kind of interpretive coordination between different conversational participants:

(32) **A**: A wolf^1 could walk in at any minute.

B: Would it$_1$ eat me?

(33) **A**: Hob thinks a ghost^1 is haunting Nob.

B: Does he think that it$_1$ is malevolent?

(34) **A**: Harvey courts a girl^1 at every convention.

B: Is she$_1$ usually a mathematician?

The exchanges in (32)-(34) are all felicitous even on the readings which assign the indefinite in each of **A**'s respective utterances narrow scope with respect to the other operator. But that does not seem to prevent those occurrences from introducing a resource that **B** can later make use of in a way that ensures that his utterances are appropriately linked to hers. Even more strikingly, the kind of intersubjective coordination secured by these means is even strong enough to support a variant of pronominal contradiction:

(35) **A**: A wolf^1 could walk in at any minute.

B: No, it$_1$ would get caught in the snare I laid outside.

(36) **A**: Hob thinks a ghost^1 is haunting Nob.

B: No, he thinks it$_1$ is haunting Cob.

(37) **A**: Harvey courts a philosopher^1 at every convention.

B: No, she$_1$ is usually a mathematician.

Remember that cases like (26) were an essential part of our linguistic motivation for endorsing the idea that certain occurrences of indefinites had singular use-conditions in the first place. But (35)-(37) appear to be entirely structurally parallel and yet are felicitous even on the readings of **A**'s utterances which seem to (if anything) tell against her having any strictly singular communicative intentions. And since we took that case to be, in some sense, among the most distinctively singular, it thus seems highly unlikely that we will be able to find a

principled reason to theoretically distinguish between the types of core cases discussed in the previous section and the ones we are presently considering.

I take the phenomenon of anaphora under operators to definitively show that we cannot always expect a speaker to have some (singular, existing) object in mind when they use an expression that introduces a novel anaphoric possibility. This shows that, among other things, we were too quick in jumping to singular communicative intentions as the right characterization of the special interpretive features exhibited by occurrences of expressions that succeed in introducing discourse referents. Instead, we need a more general characterization of those features that can also properly account for the introduction of discourse referents in these problematic cases. This more general characterization could then be used as the basis for an account of discourse entities with a wider explanatory reach. In looking for that characterization, I believe we may safely confine our attention to dependent anaphora for two reasons. First, it is the *simplest* of the counterexamples offered above. Since it arises in a purely extensional setting, we will not need to get bogged down negotiating the exact details of the richer intensional frameworks that seem necessary for handling modal and propositional attitude operators, and so there will be fewer potentially misleading distractions. But more importantly, it also is arguably the most *general* of the counterexamples offered above. It is now commonplace to model intensional operators as quantifiers over intensional entities of a given kind in semantic theory. According to that view, the type of anaphora at issue in (29) and (30) is actually entirely structurally parallel to (31), the only difference being that the quantifiers in those cases range over metaphysically or doxastically possible worlds rather than ordinary individuals. A theory of discourse entities that manages to account for (31) should thus also be able to (in principle) handle (29) and (30) if we are willing to model intensional operators in this way.[12]

[12]More accurately, we need to both be willing to model intensional operators in this way and willing to take that model seriously. That is to say, we need to understand this quantification as incurring some kind of ontological commitment to the entities that intensional operators quantify over. There is, I think, good enough linguistic reason for doing this. See e.g. Cresswell (1990) and Schlenker (2006) for arguments to that effect.

CHAPTER 4

Anaphoric Pronouns as Variables

We have just argued that discourse referents cannot be modeled as individual concepts defined in terms of a speaker's singular communicative intentions. Cases of anaphora under operators are all situations where occurrences of expressions that are not governed by singular use conditions still manage to introduce discourse referents. But this kind of theory is otherwise quite attractive since it is very well-suited to explain many different difficult phenomena (most notably, the significance of anaphoric links in cases of pronominal contradiction). Therefore, this chapter aims to modestly generalize this type of proposal and show how discourse referents can be defined in terms of a speaker's non-singular communicative intentions instead.

In §1, I discuss how the singular use-conditions governing occurrences of indefinites and proper names appear to follow from a cluster of more general use-conditions that apply to any noun phrase occurrence. If a speaker warrantedly asserts a sentence involving a proper name or (unembedded) indefinite, the evidential support of that assertion will typically concern some specific individual. In other words, a warranted assertion in these cases requires that a speaker have a witness to the truth of their claim in mind. But it is reasonably straightforward to characterize the possible witnesses to the truth of a claim involving a (possibly complex) generalized quantifier, as well. A requirement to have a witness in mind thus makes sense in those contexts, too. It appears that most occurrences of quantifier phrases introduce discourse referents whose denotations are determined by witnesses of precisely this kind. And when these general conditions are applied to cases of polyadic quantification, we

are immediately given the resources to handle dependent anaphora.

In §2, I discuss why the discourse referents introduced by certain quantifier phrases cannot be treated as being merely plurally designating. In particular, they need to also track structural dependencies with the designations of other discourse referents associated with the same licensing witness. I suggest that the best way to model this kind of structural dependence formally is in terms of Tarskian variable-meanings defined against the background of a rich state-space capable of representing every possible finite pattern of structural dependence. It follows that discourse referents can be identified with variable concepts, which are defined along similar metalinguistic lines as the proposal from the previous chapter. I then illustrate how this proposal works by giving a natural account of dependent pronominal contradiction.

In §3, I briefly discuss how my proposal relates to discourse representation theory and its various compositional reformulations, which similarly analogizes anaphoric pronouns to variables in the way that I have suggested. This work is formulated in a framework that takes a few radical departures from classical semantic architectures, making it somewhat tricky to straightforwardly characterize the contents of sentences involving anaphoric pronouns in context. These theories are also designed to automatically yield discourse referents that track witnesses directly given by conventional word meaning rather than those intended by a speaker. So while my proposal shares certain surface similarities with these earlier theories, the fact that it is explicitly formulated in terms of a speaker's communicative intentions and embedded in an otherwise standard semantic architecture means that it straightforwardly yields different judgments about the contents of specific occurrences of anaphoric pronouns.

4.1 Domains, Witnesses, and Warranted Assertability

We are now presently concerned with examples of dependent anaphora, like the one repeated below:

(31) Harvey courts a girl[1] at every convention. She$_1$ always comes to the banquet with him.

Recall that since the pronoun she may be anaphorically linked to the indefinite a girl even when it takes narrow scope with respect to the universal quantifier phrase every girl, we cannot always explain the interpretive significance of anaphora in terms of singular communicative intentions. The obvious solution to this problem is to broaden our horizons somewhat and take the fact that communicative intentions are sometimes non-singular more seriously. But in order to get a sense of how these non-singular communicative intentions can be exploited in a systematic way to explain examples like (31), we will need to consider a wider range of cases than the ones that we have been considering up to this point.

Proper names and indefinites are not the only kinds of expression that can introduce discourse referents. For example, quantifier phrases also seem to be able to establish strong interpretive connections to subsequent and grammatically distant pronouns:

(38) Most philosophers[1] don't have many friends. They$_1$ are hard to get along with.

The sequence in (38) is naturally and automatically understood as being about a particular group. In this case, that group appears to provide the domain that the quantifier most ranges over, and then later also serves as the designation of the plural pronoun they. But there are some important subtleties in this case that were not obviously present in our initial discussion of indefinites. Strictly speaking, a quantifier phrase occurrence seems capable of introducing *two* different discourse referents, which are distinguished by the fact that they designate distinct (but related) subjects of conversation:

(39) a. Most of my students[1] didn't do the reading I assigned. I forgot to remind them$_1$.

b. Most of my students[1] didn't do the reading I assigned. They$_1$ didn't pass the quiz.

c. `Most of my students`[1] `didn't do the reading I assigned. *But they`$_1$ `understood it.`

In (39a), the plural pronoun `they` is most naturally interpreted as designating all of the speaker's students in question. In (39b), on the other hand, it is instead most naturally interpreted as designating only those students who did not do the reading. I have included the infelicitous sequence in (39c) to show these are, in a sense, the only two live options. Here, the pronoun cannot be understood as designating the students who *did* do the reading, even though that group is the only one that can yield a consistent reading of the subsequent utterance and is in some sense made just as pragmatically salient by the initial utterance in the sequence as the other two. The simplest explanation of this asymmetry has the relationship between an occurrence of a quantifier phrase and the possible denotations of an anaphorically linked plural pronoun be partly grammatically determined. That is to say, the general rules governing the interpretation of anaphorically linked pronouns must somehow privilege or make reference to the options realized in (39a) and (39b), and not the one that is ruled out in (39c). In order for this kind of rule to work in full generality, you will need to be able to systematically determine what the denotational options are for occurrences of *any* given antecedent quantifier phrase. And since indefinites can be seen as a special limiting case of quantifier phrases more generally, it thus seems like we can gain deeper insight into the discourse referent-introducing aspect of their use-conditions by first paying attention to this broader class of cases.

To do this, it will be helpful to first briefly review some basic notions from the study of the semantics of quantification in natural language.[1] Assuming that we have fixed a background universe of objects U, typical quantifier phrases like `most philosophers` are interpreted by subsets of $\wp(U)$, otherwise known as *type* $\langle 1 \rangle$ *generalized quantifiers*. To continue the example, the interpretation of an occurrence of `most philosophers` would be those subsets

[1]See Peters and Westerståhl (2006) for an encyclopedic overview.

of the domain which contain more than half of the philosophers who are contextually at issue. Given a generalized quantifier Q, we say that it *lives on* a set X just in case for any Y: $Y \in Q$ iff $X \cap Y \in Q$. For most of the generalized quantifiers that interpret English quantifier phrases, we can easily define the *least* set that they live on as the intersection of all of the sets that it lives on. The least set that a quantifier lives on gives us one useful formal semantic characterization of the objects that a quantifier phrase can be used to talk about. Here, the least set a quantifier lives on is something like the domain of objects that quantifier ranges over.[2] If X is the least set that a quantifier Q lives on, a set $Y \in Q$ is a *witness set* just in case $Y \subseteq X$. We can also think of witness sets as providing another useful (and quite different) characterization of the groups of individuals that a quantifier phrase may be used to talk about. In this case, the witness sets of a quantifier are intuitively just those subsets of its domain that it positively classifies. Hence, in some sense, the quantifier can be used to talk about any one of those subsets, in addition to its domain as a whole.

With that terminological background in place, we can now shed some additional light on (39). The least set that the occurrences of the quantifier denoted by the phrase `most of my students` lives on in these examples is the set of all the speaker's students.[3] The discourse referent salient for interpreting the pronoun occurrence in (39a) appears to designate all of the students in that set. In (39b), however, the discourse referent appears to designate only the elements of a certain witness set of the quantifier – namely, the speaker's students whom both did not do the reading and failed to pass the quiz. Both of these options are admissible

[2]As Ed Keenan (p.c.) points out, this general strategy for semantically defining quantifier domains can sometimes fail if we are working with non-finite domains. For instance, the quantifier that interprets `all but finitely many numbers` lives on every cofinite set of numbers, but the intersection of those sets will be empty, and hence obviously not itself a set the quantifier lives on. So a different characterization of a quantifier's domain will ultimately be needed, but we can, for the moment, set that issue aside.

[3]More accurately, we presumably would actually want to say that it is the set of all the speaker's students among an appropriately contextually restricted group. For instance, (39a) may be uttered in a context where a certain course among many that the speaker is teaching is explicitly under discussion, and so the quantifier is naturally regarded as being implicitly restricted to that course. One of the nicer features of using least live-on sets rather than the explicitly provided restricting argument is that we may respect this aspect of a quantifier phrase's use.

since they both correspond to one of our two semantically natural characterizations of what a quantifier phrase may be used to talk about. By these lights, we would also expect the sequence in (39c) to be ruled out. Since the set of the speaker's students who did do the reading is neither the least set that the quantifier in question lives on nor one of its witness sets, it should not be a live option as the interpretation of a pronoun that is discursively linked to that quantifier.

One way of capturing the above insight is to say that occurrences of quantifier phrases are governed by the following schematic principle:

> (**QP Subjects**) On any given occasion of use, if a speaker utters an expression or complex of expressions with the intention of it being interpreted by the quantifier Q, then that speaker must also intend to talk about: (i) the individuals who make up a (minimal) set X which Q lives on, and (ii) when possible, the individuals that make up an admissible witness set of Q (relative to X).

The idea is roughly that when a speaker intends an expression to be interpreted by a quantifier, part of that includes having a domain of individuals that quantifier ranges over in mind, as well as a particular witness set of that domain. Both of these conditions seem like they can be motivated on relatively simple grounds, and so they enjoy a good deal of *a priori* plausibility. Generally speaking, being in a position to warrantedly assert a proposition involving a quantifier minimally requires an understanding of what that quantifier ranges over, and so the first part of (**QP Subjects**) seems almost rationally unavoidable. And the fact that quantifier domains are now also widely regarded as being contextually restricted in some way is further evidence for this condition.[4] After all, it seems inevitable that we will eventually need to appeal to the speaker's intentions in order to determine some of those contextual restrictions. Reflection on the preconditions of warranted assertion lends immediate support

[4]See e.g. Von Fintel (1994) and Stanley and Szabó (2000) for linguistic arguments for the claim that quantifier domains are contextually restricted.

to the second part of (**QP Subjects**), as well. In practice, the informational support for an assertion of a proposition involving a quantifier will often directly concerns some proper subregion of that quantifier's domain which is sufficient to establish the proposition's truth. And the evidential base of an assertion is always conversationally relevant. More specifically, warranted assertion of a proposition involving a quantifier will require the speaker to be in a position to adequately justify themselves if called upon to do so, which will partly involve a specification of their evidential base. But this just means that they, in fact, have some witness of the quantifier in mind that will suffice for this purpose against the background of whatever domain of quantification is at issue.

Different types of quantifier phrases then may put different constraints on what an *admissible* witness set needs to look like. In some cases, the answer is, in some sense, semantically forced on us. For instance, the quantifiers that interpret occurrences of `every` N and `no` N will both have *unique* witness sets (respectively, the least set the quantifier lives on and the empty set). In both cases, the second condition is essentially ignored.[5] But when a quantifier has many distinct witness sets, some additional constraints on which ones a speaker is allowed to have in mind are called for. Moreover, it seems like we will not be able to find a general semantic principle that will give us those constraints in every case. A familiar example drives that point home. On a given occasion of use, the quantifiers that `one or more` N`s` and `an` N designate will be exactly the same. But they appear to have different constraints on which witness sets a speaker is obligated to have in mind. When using `one or more` N`s`, a speaker is seemingly required to have a (near-)*maximal* witness set in mind – that is, a witness set that includes almost all of the elements of the quantifier's domain that also have the property in the quantifier's scope. But when a speaker uses `an` N, they are

[5]More generally, a quantifier with a (non-empty) unique witness is a principal filter. Typical examples are proper names (when viewed as quantifiers) and conjunctions thereof. Here, the least live-on set and witnesses always coincide and yield the intuitively correct use-conditions for singular terms and the lists you may construct out of them.

instead required to have a *singleton* witness set in mind.[6] Hence, an unembedded indefinite's apparently singular use conditions turn out to just be a particular way of filling out the (**QP Subjects**) schema.[7] This principle thus seems like the right place to start if we wish to generalize the theory of the previous chapter to non-singular instances of anaphoric linking, too.

This more general characterization of a quantifier phrase's use-conditions actually gets us much farther than you might have initially thought. The motivation for the (**QP Subjects**) schema came from consideration of instances of unembedded type $\langle 1 \rangle$ quantifiers. But the type of quantification at issue in cases of dependent anaphora is a bit more complicated. Again assuming a fixed background universe U, a *type* $\langle 2 \rangle$ generalized quantifier is any subset of $\wp(U \times U)$ – in other words, type $\langle 2 \rangle$ generalized quantifiers are higher-order properties of binary relations over the universe (rather than unary properties).[8] Sentences with two quantifier phrases in the same local construction are often best analyzed as actually involving type $\langle 2 \rangle$ quantification of some kind. Given two type $\langle 1 \rangle$ quantifiers Q_1 and Q_2, the *iteration* of Q_1 and Q_2 ($Q_1 \cdot Q_2$) is the type $\langle 2 \rangle$ quantifier given by $\lambda R.Q_1(\lambda x.Q_2(\lambda y.Rxy))$. Iteration is one of a few different *polyadic lifts* of monadic generalized quantifiers that are relevant to the study of multiply quantified sentences in natural language. It is, in a sense, the simplest one and corresponds to the familiar idea of assigning one quantifier determinate linear scope over another.

---

[6] Compare this view to the one presented in Schwarzschild (2002) and more recently advanced in a philosophical context by Hawthorne and Manley (2012). According to that proposal, the special scoping features of indefinites are explained by appealing to a kind of singleton domain restriction. Depending on how the details are filled out, the view presented here may be considered another variant of that proposal.

[7] There is, I suppose, a question about whether this use-conditional aspect of indefinites is better seen as being semantic or pragmatic. I am inclined to view it as a non-truth-conditional aspect of its semantics, as it seems to be conventional in a familiar minimal sense. In particular, while I think that something like the (**QP Subjects**) schema can arguably be motivated on broadly Gricean grounds, it still leaves open a non-trivial option space. The fact that indefinites have resolved those options in one particular way and another expression has resolved them in a different way is, therefore, good grounds for some claim to conventionality (cf. a similar line of reasoning regarding the special interpretive features of definites in the exchange between Szabó (2000, 2003) and Abbott (2003).)

[8] More generally, a type $\langle n \rangle$ generalized quantifier over a universe U is just some subset of $\wp(U^n)$.

Simplifying a bit, the reading of (31) that has been concerning us implicitly involves the type $\langle 2 \rangle$ quantifier $[\![\texttt{every convention}]\!] \cdot [\![\texttt{a girl}]\!]$. Now, the important thing to note is that the (**QP Subjects**) schema applies *equally well* to this polyadic quantifier, so long as we can generalize the notions of lived-on and witness sets to the polyadic case. And there are clearly no obstacles to doing so – in fact, our original definitions of these notions also make sense in a polyadic context. With this in mind, applying (**QP Subjects**) to this example is relatively straightforward. Since $[\![\texttt{every convention}]\!]$ lives on the set of contextually relevant conventions C and $[\![\texttt{a girl}]\!]$ lives on the set of contextually relevant girls G, then $[\![\texttt{every convention}]\!] \cdot [\![\texttt{a girl}]\!]$ clearly lives on the relation $C \times G$. So the only real question remaining is which witness sets the speaker is allowed to have in mind when uttering (31). Presumably, the constraints on the intended witness sets for an iterated quantifier $Q_1 \cdot Q_2$ should, in some sense, arise from the constraints on the individual quantifiers Q_1 and Q_2. The following proposal is an extremely natural way of implementing this idea:

> (**Distributive Satisfaction**) Suppose that X and Y are the collections of contextually admissible witness sets for the quantifiers Q_1 and Q_2, respectively, and that Z is the collection of contextually admissible witnesses of the quantifier $Q_1 \cdot Q_2$. Then $R \in Z$ iff (1) $\pi_1(R) \in X$, and (2) $R_a \in Y$ for each $a \in \pi_1(R)$ (where $R_a = \{b \mid aRb\}$)

In other words, R should count as an admissible witness of $Q_1 \cdot Q_2$ just in case restricting R to only its first component is an admissible witness of the wider scoping quantifier Q_1, and each set arrived at by distributively fixing the first element of R by an element of that admissible witness is itself an admissible witness of the narrow scoping quantifier Q_2. In the present context, the result will be that the admissible witnesses $R \subseteq C \times G$ will be pairings of every contextually relevant convention in C with one contextually relevant girl in G.

These witnesses supply all of the relevant information needed for interpreting pronouns that are anaphorically linked to narrowly scoped quantifiers. Consider the following illustra-

tive example:

(40) a. `My daughter`^{1} `gave a drawing`^{2} `to all of her`$_1$ `teachers`^{3}`.`

b. `They`$_3$ `each put it`$_2$ `on the wall.`

c. `They`$_3$ `each threw it`$_2$ `in the trash.`

According to the present proposal, an utterance of (40a) has to be accompanied by an intention on the part of the speaker to talk about a particular witness of a polyadic quantifier. If what we have said so far is right, these witnesses will be parings of teachers to drawings. If an utterance of (40a) is then followed by an utterance of (40b), that pairing can be used to supply the denotations of anaphorically linked pronouns: `they` is interpreted by the first coordinate, and `it` is interpreted by the second coordinate. So that witness will verify (40b) just in case the teachers hung up the drawings that the pairing associates them with. Note that this means that the truth of (40b) essentially depends on which witness the speaker had in mind. Now, suppose that the speaker's daughter gave each of her teachers *two* drawings, one of which they promptly displayed while the other they rudely disposed of. Then there will, in general, be quite a few different witnesses that verify (40a), but only one of those will *also* verify (40b). And, of course, the present proposal does not say that the speaker is *forced* to have that witness in mind. For instance, she could intend to talk about the unique verifying witness of (40c) instead. The relevant prediction made by our proposal is that the speaker may felicitously continue her initial utterance of (40a) with either (40b) or (40c). But since continuing down either of these routes requires having *different* verifying witnesses in mind, the speaker is never allowed to felicitously utter the complete sequence in (40). And this appears to be the correct thing to say.[9]

[9]I am of course *not* claiming that either continuation in (40b) or (40c) will be equally natural in all contexts. Quite the contrary – some background questions under discussion will make one clearly better than the other. Hence, if you are discussing how impressive your daughter's artwork is, then a witness that verifies (40b) is obviously more contextually appropriate. And if you are talking about how callous your daughter's teachers can be, then a witness that verifies (40c) will instead be the one that is contextually preferred. My only claim is that neither reading is ruled out on strictly grammatical grounds.

A similar explanation can also be given for (31), repeated here with one necessary amendment:

(31′) `Harvey courts a girl`^{1} `at every convention`^{2}`. She`$_1$ `always`$_2$ `comes to the banquet with him.`

Since Lewis (1975), it has become standard to treat adverbial quantifiers as ranging over a contextually supplied case base. Following Brasoveanu (2008, 2010), we now suppose that often that case base is supplied by an anaphoric link to an earlier suitable quantifier phrase. The explanation given for (40) now carries over essentially unchanged. A speaker uttering the first sentence of (31) needs to have a particular admissible witness of the iterated quantifier in mind. If she is speaking truthfully and her assertion is warranted, it will be a paring that maps each of the conventions that Harvey attended to some girl he courted at that convention. The quantifier `always` is then interpreted by the first component of that witness, and the pronoun `she` is interpreted by the second. It follows that the second sentence will be true just in case each element $\langle \mathbf{c}, \mathbf{g} \rangle$ of that witness is such that Harvey brought $\mathbf{g}$ to the banquet of $\mathbf{c}$.

4.2 Discourse Referents as Variable Concepts

Assuming that what we have said so far is right and that (**QP Subjects**) and (**Distributive Satisfaction**) are the interpretive principles at work in cases of dependent anaphora, what is the resulting view of discourse referents that we are left with? At one level of description, the answer is relatively straightforward – they may still be identified with concepts whose possibly plural denotations are determined by the communicative intentions of the discourse participants. According to this view, the entities that discourses are organized around are derived from the quantifier domains and witnesses that a speaker has in mind when she makes a warranted assertion. Moreover, these entities are presented to an interpreter in precisely that way – namely, as the very ones the speaker had in mind when using this or

that expression. But while I think this basic idea is more or less correct, putting things in exactly those terms oversimplifies the situation in a conceptually misleading way. In particular, plural denotation *on its own* is insufficient to capture the phenomena at issue here adequately. The easiest way to appreciate this general point is just to note that the most natural reading of (say) (40b) is, in fact, *stronger* than the above story seems to predict. Its truth-conditions are not *just* that the teachers hung the drawings on the wall – instead, it is something like each teacher hung *the drawing that was given to him* on the wall. And it is very unclear how we could recover something like this result if all we are given in interpretation is a handful of plurally designated teachers and drawings.

Of course, one might maintain that the above stronger reading is due to some kind of pragmatic enrichment and should not strictly be part of the truth-conditions delivered by interpreting pronouns in this way. In order to ward off thinking along those lines, we will need to turn to a slightly more complex technical problem that is clearly of a grammatical (i.e., non-pragmatic) nature. To start, note that the witness sets of a given polyadic quantifier can be visualized as a matrix. As an example, suppose that we are interpreting (40a) and the speaker intends to talk about a witness set which respectively pairs the teachers $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_3$ with the drawings $\mathbf{d}_1$, $\mathbf{d}_2$, and $\mathbf{d}_3$. Then that witness set looks something like this:

$$\begin{array}{ll} \mathbf{t}_1 & \mathbf{d}_1 \\ \mathbf{t}_2 & \mathbf{d}_2 \\ \mathbf{t}_3 & \mathbf{d}_3 \end{array}$$

Thinking about witness sets in this way makes it clear that actually, there are two relevant kinds of information that they carry. The columns of the matrix encode information about the objects designated by a given discourse referent, and the rows encode information about certain functional dependencies that hold between them. Talking about discourse entities in the way suggested above makes it sound as though the information carried by the individual columns completely exhausts what is interpretively pertinent. But the information carried

by the rows is equally important. The easiest way to see that is to contrast (31) with the sequences below:

(41) a. `Harvey courts a girl`^{1} `at every convention. She`$_1$ `is very pretty.`

b. `Harvey courts a girl`^{1} `at every convention. They`$_1$ `are very pretty.`

Unlike (31), (41a)'s only admissible reading grants the indefinite scope over the universal quantifier. And the sequence in (41b) appears to resolve the scope ambiguity in the other direction. The addition of an adverbial quantifier thus seems to have an important role in licensing the use of a *singular* pronoun to be interpreted by what appears to be a *plurally designating* discourse referent. The main technical challenge in this area is explaining why the presence of additional quantifiers in examples like (31) and (40) make this possible.

The solution to this problem requires that we additionally appeal to the information carried by the individual rows of a polyadic quantifier's witness sets.[10] The rough idea is that in the cases where we distributively quantify over one of the columns of a polyadic quantifier's witness set (as we are apparently doing in (31) and (40)), the semantic singularity condition of a singular pronoun like `she` or `it` that is interpreted by another column of that witness set is only checked against certain *subsections* of that column. And those subsections are determined by appealing to the *rows* of the witness set in question. In particular, they are the subsections obtained by pointwise fixing single elements of the column that you are distributively quantifying over. In (31) and (40), the interpretive principle (**Distributive Satisfaction**) will then guarantee that the singularity condition is met, hence explaining why the singular pronoun is admissible in those cases. In the absence of a distributive operator of some kind, the singularity condition is instead checked against the entire column that the pronoun is interpreted by. This explains why (41a), if well-formed, must truth-conditionally assign the indefinite wide scope. If discourse referents were treated as *merely*

[10]This solution is derived from work on plural discourse reference in the kind of broadly semantic framework that we have been assuming here. See e.g. Elworthy (1995), Krifka (1996), van den Berg (1996), Nouwen (2007), and Brasoveanu (2008, 2010).

plurally designating, this kind of explanation would be impossible. In addition, a discourse referent will occasionally need to keep track of the structural dependencies that hold between its denotations and the denotations of the other discourse referents that were derived from the same intended witness or domain.

At this point, you may recall that in an earlier chapter, we argued that the meanings of variables in a Tarskian framework served more or less the same kind of theoretical role that we are now seeing discourse referents play. In other words, they were concepts that were capable of tracking *both* the values that particular variables individually range over *and* the acceptable patterns of valuation that an entire system of variables was allowed to assume simultaneously. When this idea was presented in that earlier setting, I informally suggested that concepts of this kind were reasonably commonplace. For instance, in thinking about *your students* and *their parents*, it seems like you are thinking about two collections of people in such a way that additionally involves implicit structural relationships between them. However, the basis for that claim was primarily intuitive. We are now in possession of a much stronger argument – in particular, it seems like the proper semantics of anaphoric pronouns straightforwardly requires concepts of precisely this kind.[11]

We thus seem to have arrived at the view that discourse referents are roughly the same kind of thing as Tarskian variable-meanings. According to that theory, the interpretation of a variable can be abstractly modeled as a function from possible states to the values that the variable assumes in those states. The possible structural dependencies between distinct variables can then be captured by restricting the values that those variables may simultaneously assume across the entire state-space. However, we only directly worked with one concrete realization of this idea – namely, the Tarskian interpretation of variables in in first-order predicate logic. In that context, states are modeled as variable assignments

[11]Notice that a Fregean account of this kind of structural dependence is entirely ruled out by the fact that was noted at the end of Chapter 1. In particular, discourse is obviously interpreted incrementally in (at least) the very weak sense that it is interpreted utterance by utterance. And, as we have already seen, anaphoric dependencies can stretch between these utterance boundaries. But a Fregean strategy would be flatly incompatible even with this weak form of incrementality.

(alternatively, ω-sequences), and this way of doing things places no structural restrictions on the values that any two variables are allowed to simultaneously assume. So in what follows, we will need a much more general state-space to capture the phenomena we are presently discussing. In particular, it will need to be capable of representing every (finite) pattern of structural dependence that may emerge throughout a conversation. In other words, the state-space we work with will be such that for any n-ary relation R, there will be n variable-meanings $v_{i_1}, \ldots, v_{i_n}$ such that $\{\langle a_1, \ldots, a_n\rangle \mid \exists\sigma.v_{i_1}(\sigma) = a_1 \;\&\ldots\&\; v_{i_n}(\sigma) = a_n\} = R$. Explicitly constructing a space of this kind is a straightforward but somewhat technical exercise, and the exact details of how the space is concretely realized will not matter much in what follows. So we will just omit those details here and instead proceed on the basis of the more abstract characterization of the space just given.[12] A *variable concept* can then be identified with a function from possible worlds to variable-meanings defined in terms of this background state-space.[13]

My suggestion is that we model discourse referents as variable concepts in the above sense, in more or less the same way that the Stalnakerian theory of the previous chapter attempted to model them as ordinary individual concepts. In order to see how this can be done, let us, for the moment, focus just on those discourse referents that arise from a speaker's evidential base. According to (**QP subjects**), a speaker that warrantedly utters a (multiply-)quantified sentence S whose analysis involves a type $\langle n\rangle$ quantifier will generally speaking have an n-ary relation in mind to serve as a witness. So it follows more or less immediately that a certain metalinguistically defined *relational concept* will be available to an interpreter to make use of in that circumstance, which we may approximate as *the witness that the speaker had in mind when uttering* S. An interpreter can further descriptively isolate the components of this intended witness by noting their associations to distinct occurrences

[12]See Appendix B for a sketch of how to construct a space with the desired properties.

[13]Obviously, these may be alternatively thought of as binary functions from a world and a variable-state to an individual.

of expressions. For example:

(42) `At least three students`[1] `completed exactly half of the questions`[2]`.`
`But they`$_1$ `answered them`$_2$ `correctly.`

Simplifying somewhat, when a speaker utters the first sentence of (42), the proposition that they express will ordinarily involve a type $\langle 2 \rangle$ quantifier that arises from iterating the type $\langle 1 \rangle$ quantifiers $[\![$`at least three students`$]\!]$ and $[\![$`exactly half of the questions`$]\!]$. Following (**Distributive Satisfaction**), a contextually appropriate witness of this quantifier will be a relation R such that $\pi_1(R)$ is a contextually appropriate witness of $[\![$`at least three students`$]\!]$ and that each set arrived at by distributively fixing the first element of R is a contextually appropriate witness of $[\![$`exactly half of the questions`$]\!]$. If the speaker is warranted and speaking truthfully, it follows that they will have a relation in mind that pairs each of three contextually appropriate students with the half of the contextually appropriate questions they managed to complete. Since the constraints on the first component of this witness were derived from the rules governing the occurrence of the expression `at least three students`, that component can be descriptively isolated crudely as the witnessing things the speaker had in mind when uttering `at least three students`. And since the constraints on the second component of this witness were derived in part from the rules governing occurrences of the expression `exactly half of the questions`, that component of the witness can similarly be descriptively isolated crudely as the witnessing things the speaker had in mind when uttering `exactly half of the questions`.

The above information suffices to formulate a system of n variable concepts, which can then be taken to model the discourse referents introduced by their associated expression occurrences. In particular, for each epistemically possible world w, the variable-meanings $v_{i_1}, \ldots, v_{i_n}$ these concepts yield in w will jointly represent the intended witness in w. To continue our example, the variable concepts at work in (42) will jointly represent the intended association of students to problems in the actual world. These concepts are then used to

fix the contents of subsequent anaphorically linked pronouns as their associated variable-meanings. This suggests that we can crudely represent the semantic structure of (42) by something like the following:

(42′) $[\text{THREE } \mathtt{v}_1 : \textsc{student}\ \mathtt{v}_1][\text{HALF } \mathtt{v}_2 : \textsc{question}\ \mathtt{v}_2]\ \textsc{completed}\ \mathtt{v}_1\mathtt{v}_2\ \wedge$
$\textsc{correctlyAnswer}(\mathtt{v}_3, \mathtt{v}_4)$

To evaluate this formula, the variables $\mathtt{v}_3$ and $\mathtt{v}_4$ need be interpreted by variable-meanings v_3 and v_4 which together jointly represent the witness the speaker had in mind when uttering (42), with v_3 associated to the occurrence of the expression `at least three students` and v_4 associated to the occurrence of the expression `exactly half of the questions`. Note that the second sentence of (42) has been symbolized as an open formula, which directly reflects the fact that the pronouns involved are natural language analogs of variables in first-order languages, and so are to be interpreted by the same kind of thing. Following standard practices, we say that (42′) is true simpliciter just in case it is true with respect to every possible state of the variables in question. This results in the variables $\mathtt{v}_3$ and $\mathtt{v}_4$ being treated as though they were implicitly universally quantified. But since they are *restricted* variables, the result is actually the desired reading – (42′) is true just in case at least three of the contextually relevant students completed exactly half of the contextually relevant questions, and the intended witnessing students correctly answered their distributively associated intended witnessing questions.[14]

Stepping back for a moment, we should briefly review how we got to this point and discuss how this theory addresses some of the central concerns of discourse semantics. As noted in the previous chapter, a discourse's subject matter provides a crude but effective way of characterizing cross-speaker coordination. If an episode of communication between two or more language users is going to be entirely successful, they must all be talking about *the same things*. Given the importance of this kind of coordination, we should expect at least a

[14]See Appendix B for an explicit statement of the semantics of the formal language used to represent these truth-conditions, together with a more detailed discussion of how it can be applied in similar examples.

few different interpretive mechanisms for generally securing it. And in fact, it seems like the semantic significance of anaphoric co-indexation is more or less directly geared towards this purpose. The basic idea behind the Stalnakerian theory presented was that an interpreter can *guarantee* a kind of referential alignment by explicitly invoking the referential intentions of the speaker when resolving the contents of certain expressions. More explicitly, a cheap way for an interpreter to ensure that she is referring to the same thing that a speaker is referring to is by using a content that can be approximately glossed as *the object that the speaker intended to refer to*. According to this outlook, whenever two linguistic items are co-indexed, we partly signal that an interpreter may use something like the above method to ensure that all co-indexed expressions designate the same subject of conversation. Each different index is then supposed to correspond to a *different* descriptive content of this kind. However, as we have now also noted, co-indexation also seems appropriate in cases that cannot be easily captured by a speaker's intention to refer to a singular object of conversational attention. The present proposal deploys the same general explanatory strategy but adjusts for this problem by employing a speaker's intentions to refer to a witness of a (possibly complex) quantifier instead.

It follows that the present theory will explain puzzle cases in more or less the same way that the singular version presented in the previous chapter did. Let us reconsider the pronominal contradiction in this new light as a helpful illustration. Our earlier example of this phenomenon is repeated here:

(27) **A**: `John`^{1} `just threw out an old broken typewriter`^{2}`.`

B: `It`$_2$ `wasn't a broken typewriter -- it`$_2$ `was a brand new computer. And he`$_1$ `did not throw it`$_2$ `out -- he`$_1$ `donated it`$_2$ `to charity.`

Our initial idea was that the occurrence of the expression `an old broken typewriter` needed to be accompanied by a certain referential intention on the part of the speaker to talk about some specific object. Assuming that the conversation is proceeding normally,

the object that the speaker intends to refer to will satisfy any conditions that the indefinite ascribes to it. Cases of pronominal contradiction are situations where the conversational context may be defective, in the sense that the speaker may have *misclassified* the object that he intended to refer to. But note that the individual concepts that we earlier identified discourse referents with do not, in fact, *require* that they be correctly classified. In other words, there may still be an object that the speaker intended to refer to when uttering the indefinite `an old broken typewriter` without there being an old broken typewriter that the speaker intended to refer to when uttering the indefinite `an old broken typewriter`. A conversational partner may thus still be able to *correct* the speaker using a co-indexed pronoun even when that object fails to satisfy essentially all of the properties explicitly ascribed to it, assuming that she is also in a position to independently identify the object in question.

We may now straightforwardly generalize this explanatory story to more complicated cases, given our current theory. For example, consider the following possible variation of (42):

(43) **A**: `Every student`^{1} `correctly answered exactly half of the questions`^{2}

B: `No, they`$_1$ `only received partial credit on some of those`$_2$

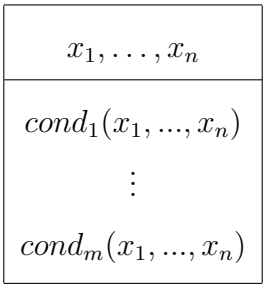
Supposing that **A**'s assertion is sufficiently warranted, (**QP Subjects**) and (**Distributive Satisfaction**) together predict that he has a witness of a complex quantifier in mind when making his assertion, which will be some pairing of students to questions. The individual components of this witness are then independently isolable as being associated with the occurrences of the expressions `every student` and `exactly half of the questions`, respectively, which can then be modeled in context as distinct Tarskian variable-meanings. In other words, the fact that **A** has a particular witness of the quantifier in mind when uttering the entire sentences grounds the claim that he has a particular variable-meanings in mind when he utters `every student` and `exactly half of the questions`. When the

conversation is proceeding normally, the objects that those variable-meanings range over are assumed to satisfy all of the conditions explicitly ascribed to them by the utterance. If v_1 is the variable-meaning associated to `every student` and v_2 is the variable-meaning associated to `exactly half of the questions`, then the default assumption in this case is that (aside from the cardinality constraints) v_1 and v_2 only simultaneously assume values where v_1 is a student and v_2 is a question that v_2 answered correctly. But of course, **A** may be mistaken about one or more of these features. The whole exchange in (43) is best understood as involving a misunderstanding of this kind. In particular, it is natural to suppose that **A** has succeeded in isolating variable-meanings that satisfy the relevant cardinality constraints and most of the ascribed conditions. In other words, he intends to refer to all of the contextually relevant students and some questions that they were assigned in such a way that structurally associates exactly half of the total number of assigned questions to each of the students at issue. But he mistakenly believes that those questions are the ones the students answered correctly. When **B** corrects him, she uses a co-indexed pronoun whose designation is fixed by using a descriptive content that refers to **A**'s structured plural referential intentions and so is guaranteed to isolate the same variable-meaning as him. But knowing better, she can then inform him that the student's performance on the exam was even worse than he initially supposed.

The result is a theory that assumes that language users can more or less automatically secure a much more robust form of interpersonal coordination than the theory presented in the previous chapter predicted. In directly coordinating on a more general class of witnesses, they are guaranteed to agree not only on the things that their conversation is about but also on certain *relations* that hold between those things. And while this guarantee admittedly only holds at some very high level of abstraction, we have already seen that it still seems powerful enough to shed light on a wide array of quite complicated linguistic phenomena.

4.3 Comparison with Other Proposals

We have just argued that anaphoric pronouns function like the variables of first-order languages and suggested that their contents in context be formally modeled with the same kind of meanings. Our proposal obviously owes a great deal to earlier work on discourse reference, which analogizes pronouns to variables in a somewhat similar way. So in order to appreciate what is novel about the present account, we will need to quickly summarize the main points of those foundational contributions first. For our purposes, the most important early example is Kamp (1981)'s *Discourse Representation Theory*.[15] It takes the form of an explicit algorithm that takes sentences in some language to be interpreted and translates them into *discourse representation structures*. In the most straightforward cases, a discourse representation structure consists of two main elements: a universe of *discourse markers* and a set of *conditions* imposed on that universe. Individual discourse markers are assumed to be drawn from some stock of variables, and conditions are then naturally identified with open formulae that may include those variables. Typically, a DRS is given a graphical representation like the following:



Kamp is quite clear that he regards DRSs as possessing some degree of psychological reality and should hence be regarded as cognitive in nature. The remarks he offers on how to interpret his semantic theory more-or-less directly correspond to the model of interpretation sketched in Fodor (1975). More specifically, this outlook presupposes something like a linguistic model for understanding the representational capacities of the mind. The inten-

[15]See also Heim (1982) for an early proposal along very similar lines.

tional states of a rational agent are assumed to be explained by reference to certain *mental representations* those agents possess, where those representations are then understood as something like discrete structures in an internal language of thought. Importantly, this internal language admits a form of syntactic description (at least in principle, if not in practice), allowing individual structures to be distinguished without appeal to their representational properties. Against this background, communication is thought of as a more-or-less translational exercise. In particular, an agent will have a specific private mental representation that codes some information they would like to make public. Their knowledge of some language equips them with the capacity to map that internal representation into some public symbol (which is standardly some sequence of articulated sounds, though, in principle, any physically detectable medium can do just as well.) The interpreter's competence with the same language then allows her to construct her own private mental representation from that physically detectable signal. The DRS construction algorithm is intended to be an abstract characterization of the capacity exercised by the interpreter in this general scenario, with DRSs corresponding to something like the private mental representations constructed during this process. So, according to this interpretation of the theory, discourse markers are the variable-like mental symbols of an interpreter that are deployed to construct relationally linked repositories of singular information. These mental symbols are then understood as being individuated along approximately syntactic lines.

Unsurprisingly, this proposal faces a few conceptual and technical challenges. On the conceptual end, it is obviously committed to a reasonably strong form of the representational theory of mind. And while there ultimately might be good reasons to endorse that theory, it is generally better to avoid accruing any specific commitments to proposals which at least appear to answer very different kinds of questions. So if our goal is *just* to characterize certain high-level interpretive facts about how discourses are organized, other things being equal, a proposal that can remain neutral on this point would be preferable. This specific theoretical commitment also puts Kamp's discourse representation theory quite out of step

with other mainstream work in semantic analysis. As we have already noted (at least in its original formulation), it takes the form of a translation algorithm into a specialized logical language. That translation is then given a model-theoretic interpretation, as usual, thereby also (indirectly) assigning a model-theoretic interpretation to the source text that served as the original input to the algorithm. While (following Montague (1973)) this is a more-or-less standard technique in formal semantics, the intermediate stage of translation is also generally regarded as a mere technical convenience that is, in principle, eliminable. Methodologically speaking, the theoretical gold standard is a direct and compositional model-theoretic interpretation of the input natural language structures. Assuming that the model-theoretic interpretation of the target language is compositional and that the translation algorithm itself works in a bottom-up direction, arriving at a direct model-theoretic interpretation of input structures is completely straightforward – merely taking the composition of the two procedures will usually suffice. The issue is that the original DRT translation algorithm essentially works in a *top-down* direction. Global facts about the parse of an entire text often directly influence its decisions about how to translate individual words.[16] The result is a theory that appears to *essentially* involve an intermediate step of translation.[17]

[16]Without going into too much detail, here is a basic illustration of this phenomenon: a bottom-up translation algorithm works in a roughly word-by-word manner, in the sense that it first specifies a translation for each word and then derives the translation for larger expressions from the translations of its components parts. For instance, a roughly Montagovian translation algorithm might assign the lambda-calculus expression $\lambda Q.\exists(\text{WOMAN})(Q)$ as the translation for the indefinite expression `a woman`, which is itself derived from the individual translations of `a` and `woman` ($\lambda P.\lambda Q.\exists(P)(Q)$ and WOMAN, respectively.) The top-down algorithm at work in classical DRT, on the other hand, does not assign an independent translation to indefinites. Instead, the interpretive significance of an expression like `a woman` is to add a new discourse marker x to an accessible DRS's universe and add the formula $\text{WOMAN}(x)$ its list of conditions. Importantly, which variable is chosen to be the new discourse marker depends on what markers have already been used. So the same source indefinite expression may ultimately be translated by different target expressions depending on global contextual considerations.

[17]This is, of course, regarded by Kamp as unavoidable. His position is that the subtleties at play in discourse interpretation require a translational intermediary. But any plausible argument to that effect requires considerations that seem to go well beyond the strict confines of the semantics of anaphora proper. See, e.g., Kamp (1990), which partly attempts to make a case for this claim by appealing to the typical battery of semantic considerations in favor of the representational theory of mind (e.g., the semantics of propositional attitude reports and the problem of logical omniscience.) While anaphora can often play a role in arguments of this kind, it is not clear that it adds any essentially new considerations.

As a reaction to some of these problems, various compositional and (supposedly) non-representational reformulations of Discourse Representation Theory have been proposed.[18] Most of these revised proposals are built (to varying degrees) on top of the semantic insights of Groenendijk and Stokhof (1991). In essence, these proposals can all be described as replacing the cognitively understood and syntactically individuated discourse markers of Kamp with entities that are individuated by a set of much more abstract criteria. Roughly speaking, these new entities are meant to serve as syntax-independent abstract informational containers, modeled on the cells in memory of an idealized computer.[19] They thus supposedly avoid commitment to psychological entities of a certain kind at the cost of admitting a class of (for lack of a better term) informational entities. These informational entities then allow these theories to describe the dynamics of information flow in an unfolding conversation at a higher level of description than was possible in Kamp's version of the theory. In particular, while these theories are designed to explain facts about certain interactions that agents may have with one another, they also apparently do not need to make any substantive assumptions about the underlying cognitive architectures of the agents under consideration.

However, both classical discourse representation theory and its compositional reformulations unfortunately tend to run together two distinct issues. On the one hand, they are explicitly concerned with providing a theory of discourse referents, understood as a way of clustering information about a subject of conversation. But semantics is then understood to

[18]To some extent, even Kamp eventually caved to pressure on the question of compositionality. See, for instance, the form that DRT takes in Van Eijck and Kamp (1997). Of course, his reasons for favoring a compositional formulation of the theory differ from the worries outlined above. The issue is that a top-down DRS construction algorithm does not directly yield a psychologically realistic theory of sentence processing, which, as we have already noted, proceeds in a roughly word-by-word manner. A bottom-up algorithm fairs much better on this front and so is ultimately preferable. But importantly, said bottom-up algorithm is still seen as characterizing a real psychological process – namely, the *incremental* construction of a mental representation as a text is being read. Hence, we can see two *very* different kinds of motivations for the principle of compositionality at work. Some see compositionality as a way of securing an abstract theory of interpretation that avoids taking on any commitments to overtly psychological claims. But others see it as a way to satisfy some of the additional theoretical constraints that an overtly psychological theory of interpretation is subject to.

[19]See e.g. the *pegs* of Landman (1986), the *referents* of Vermeulen (1995), or the *pigeon-holes* of Muskens (1996).

also be responsible for providing a characterization of the mechanisms responsible for maintaining and updating that cluster of information over time. For instance, the representational approach assumes that an interpreter constructs a mental file where vehicles carrying that information are explicitly inscribed and where new vehicles carrying new information may be progressively added as conversation proceeds. And the non-representational approach instead just assumes that abstract informational containers accrue more and more restrictive conditions on their possible occupants while attempting to remain neutral on the lower level implementational details that these facts supervene on. But on either view, semantic values are assumed to be instructions for carrying some of these processes out. The meaning of a sentence is identified with its capacity to alter the informational state of an interpreter (sometimes called its *context change potential*). In a slogan, meanings are *programs* and not *propositions*.

This further change in perspective makes applying classical semantic concepts like content and subject matter substantially more difficult. In order to appreciate the problem, let us consider a caricature of how these theories handle cases of (singular) anaphora. The informational states of an interpreter are modeled as sets of *referential possibilities*, which can, for simplicity, be modeled as partial functions from natural numbers (i.e., a set of discourse markers) to the entities in the domain of discourse. The presence of a referential possibility i such that $i(n) = a$ in an agent's information state then roughly signifies that for all that agent knows, the actual value of the discourse referent associated with the marker n is a. Sentence meanings are modeled as functions from information states to information states that can perform one of two actions:

(44) a. `A graduate student went to the convenience store.`

b. `He bought cigarettes.`

The meaning of (44a) is an extensive update that introduces a new discourse referent. It takes an information state I, selects an unused discourse marker n, and extends every partial

function $i \in I$ to now be defined on n with a suitable verifying witness (i.e., a graduate student who went to whatever convenience store is at issue). The meaning of (44b) is an eliminative update that reduces uncertainty about the values of the discourse referent already introduced. If n is the marker associated with the antecedent of the anaphoric pronoun `he`, it will take an information state I and remove any $i \in I$ such that $i(n)$ did not buy cigarettes.

But assigning the sentences in (44) updates of this kind makes it extremely difficult to characterize the content that they actually communicate in context in a wholly satisfying way. Supposing that (44a) and (44b) were uttered in sequence, these theories standardly assign them existential truth-conditions that are crudely representable as follows:

(44′) a. $\exists x(\textsc{GradStudent}(x) \wedge \textsc{WentToStore}(x))$

b. $\exists x(\textsc{GradStudent}(x) \wedge \textsc{WentToStore}(x) \wedge \textsc{BoughtCigarettes}(x))$

But these truth-conditions alone seem to completely mischaracterize the proposition expressed by (44b), at least. To appreciate this point, consider the following slight variation of (44):

(45) **A**: `A graduate student`^{1} `went to the convenience store. He`$_1$ `bought cigarettes.`

B: `No, he`$_1$ `didn't buy cigarettes. He`$_1$ `bought beer.`

B's utterance is naturally and automatically interpreted as a rejection of the second of **A**'s two claims. But if these utterances are all taken to express existential propositions along the lines suggested above, their respective utterances would be (to use some antiquated terminology) subcontrary, not contradictory. These contents are, therefore, alone insufficient to adequately characterize the nature of this dispute since their utterances could be simultaneously true.

The coordination between these two speakers is much stronger in the sense that they are clearly arguing about the features of a particular object of joint attentional focus. In

other words, we would like the content of **B**'s utterance to somehow reflect the fact that it is a straightforward negation of the content of **A**'s. The most straightforward way to deliver this result would be to construe their dispute as instead one about the value of some particular discourse referent – namely, the one introduced by the occurrence of the indefinite `a graduate student`. But this would make the contents of their utterances singular, not existential, which is not apparent from the updates the theories assign to the sentences at issue. So if these theories are to be interpreted as somehow implicitly trafficking in singular contents of this kind, the emphasis on the informational change brought about to an interpreter's information state greatly obscures that fact. And even if we grant that it is possible to interpret these theories in this way, it is still highly unclear what these theories think the values of discourse referents, in fact, are. More specifically, they are designed to automatically compute the constraints on the denotations of possible discourse referents from the semantics of the conditions directly ascribed to them. This makes it impossible for the value of a discourse referent associated with an occurrence of an indefinite like `a graduate student` to fail to be a graduate student. But there are clearly cases of corrective discourse that require this to be the case:

(46) **A**: `A graduate student`[1] `went to the convenience store to buy cigarettes.`

B: `Actually, he`$_1$ `is an undergraduate.`

The issue is that the theories we have been discussing treat discourse referents as though they are tracking the *semantic referents* of indefinite expressions (insofar as that makes sense). But this is a case where it is clear that they need to be understood as instead tracking the *speaker referents*. In isolating the theory of discourse referents from the rest of the semantic apparatus normally associated with it and emphasizing the role of speaker reference, my proposal is thus much better positioned to yield the content judgments of the required kind.

CHAPTER 5

Content on Demand

It goes without saying that pronouns are context-sensitive expressions *par excellence*, and we have just sketched a view according to which the contents of anaphoric occurrences of pronouns are fixed in context by the discourse referents associated with their antecedents. In this final chapter, we tie up some loose ends and briefly illustrate how this proposal can be integrated into a standard framework for thinking about context sensitivity in natural language semantics more generally. As it turns out, this account also has the resources to explain other types of pronoun occurrence, and so we may treat all pronoun occurrences as being governed by the same general interpretive principle. But this principle also looks like it may come into conflict with certain other important theoretical desiderata. Reconciling this conflict requires paying closer attention to some of the oft-ignored minutiae of how interpretation normally unfolds.

In §1, I introduce the basics of Kaplan's influential theory of indexicals, which is the model that most work on context-sensitivity in natural language is now based on. Its key insight is the distinction between *character* and *content*. Roughly speaking, the content of a sentence in context is what the sentence has been used to say, which may vary from occasion to occasion. The character of a sentence, on the other hand, is the context-invariant semantic rule that can be used to determine the sentence's content on a given occasion of use. Formally speaking, that means that evaluating the truth of a sentence requires a specification of two distinct groups of parameters – a content-fixing context of utterance and a content-assessing circumstance of evaluation. What sets Kaplan's framework apart

from other multiply-indexed theories is its insistence on a principle that I call *strong compositionality*, which requires the content of a complex expression to be derivable from the contents of its parts. Weaker forms of compositionality only hold at the level of character, which undermines the conceptual importance of the character/content distinction in the first place.

In §2, I review a recent argument against the Kaplanian framework that takes the semantics of quantification as its starting point. This argument claims that the ability of a distant quantifier phrase to bind a grammatically unrelated pronoun furnishes the resources needed for a counterexample to strong compositionality. While I believe that this argument ultimately fails, it does manage to highlight a tension that is worth paying closer attention to. In particular, it looks like it is impossible to give an unambiguous treatment of all pronoun occurrences while also holding on to strong compositionality and maintaining that no context shifts occur over the course of the evaluation of a sentence. Since strong compositionality is customarily understood as partly including a prohibition on context shifting operators of this kind, the popular responses to this problem choose to either deviate from the Kaplanian framework or insist that there is a special and quite radical kind of ambiguity that pronoun occurrences participate in. But by separating strong compositionality from the idea of context-shifting, I open the door to a different kind of response that has yet to be pursued.

In §3, I sketch the first ingredient of this new solution by stating a univocal context-invariant rule that governs the interpretation of all pronoun occurrences. Using the basic account presented in the previous chapter as a foundation, I show how it is possible to treat all pronoun occurrences on the model of paradigmatic anaphoric ones. In particular, deictic occurrences can be interpreted by a discourse referent corresponding to a demonstrative antecedent, and bound occurrences can be interpreted by a discourse referent corresponding to a domain-restricting nominal antecedent. But treating all pronoun occurrences as anaphoric in this way obviously requires evaluating them against the background of contexts that al-

ready contain the discourse referents corresponding to their antecedents. So it looks like expressions with the ability to introduce novel anaphoric possibilities need to be treated as context shifters of a certain kind that update contexts with suitable discourse referents.

In §4, I conclude by showing how context shifts of this kind can be accommodated without violating strong compositionality. The idea, very roughly, is that if a segment of text is interpreted incrementality, there will be room for context shifts to occur in between each episode of retrieving an expression's content and integrating it into a running estimate of the content of the overall text in a strongly compositional fashion. In other words, the ordinary formulation of strong compositionality as a prohibition on context shifting can be replaced with the (strictly weaker) requirement that the content of an expression be retrievable *on demand*, as soon as it is encountered.

5.1 Character, Content, and Strong Compositionality

The conceptual framework presented in Kaplan (1989a,b) is still, to this day, the starting point for most theorizing about the role that context should play in semantics. No doubt much of its staying power is due to the extremely elegant way it explains a certain initially puzzling fact about English indexicals like `I`, `here`, and `now`. On the one hand, these words clearly seem to have some important descriptive dimension of meaning. Ask someone what the word `I` means, for instance, and you are likely to be given a description like `the person who is presently speaking` as their initial response. But it would also be incorrect to suppose that English indexicals are in any way *synonymous* with these readily forthcoming descriptions, since they are typically used to make quite distinct claims:

(47) a. `I exist.`

b. `The person who is presently speaking exists.`

Note that (47a) (as uttered by a speaker s) and (47b) appear to have very different truth-

conditions – (47a) is false in any world where s does not exist, of which there are presumably many. And in many of those worlds where s fails to exist, (47b) still may be true, as the description `the person who is presently speaking` may be satisfied by someone other than s. Hence, they are not freely interchangeable in the way that true synonymy requires.

In more theory-laden terms, it appears better to say that the associated description serves only to *fix the referent* of an indexical on a given occasion of use. But this also seems to leave us in an awkward position. The paradigm cases of reference fixation are when something like a proper name is first introduced into a speech community. In those cases, we use whatever mechanism we have at hand (a description, overt demonstration, or the like) to establish a new means for talking about a certain object. But importantly, in these cases, the mechanism we use to fix the expression's referent is in no way a systematic part of its meaning. Once it has served its linguistic purpose, it can be quickly discarded. This stands in stark contrast to English indexicals, for which the reference-fixing mechanism appears to be their most semantically central aspect. The conceptual challenge is to design a semantic architecture that can naturally accommodate the uniform and systematic reference-fixing nature of an indexical's meaning.

It is widely assumed that the truth of a sentence as used on an occasion may depend on a wide array of distinct semantic parameters – times, worlds, locations, and so on. Kaplan's key insight was the observation that a sentence's sensitivity to a given parameter manifests itself in one of two different ways. One way rests on the common-sense idea that a sentence is used on a given occasion to say something. The truth of what has been said can then itself be seen to vary along one or more axes. Hence, a semantic parameter can supply some of the essential information for evaluating the truth of what a sentence says on a given occasion of use. The other way rests on another common-sense idea, according to which one and the same sentence may be used on different occasions to say different things. So a semantic parameter may instead play a role in determining what a sentence has been used to say on an occasion of use.

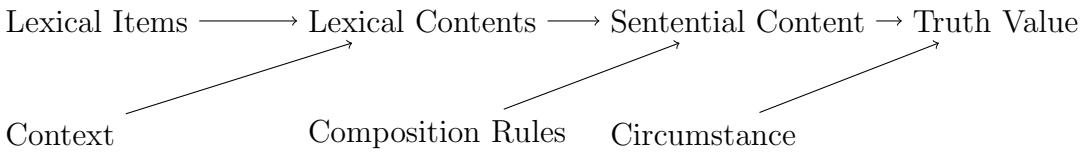
Formally speaking, Kaplan implemented this distinction in his Logic of Demonstratives by evaluating the truth of a sentence in a model with respect to two different sequences of parameters: the *context of utterance* and the *circumstance of evaluation*. These two factors model the two different ways that a sentence's truth is variable with respect to a given parameter. Intuitively, the context of utterance models an actual or possible situation in which a sentence could be uttered. Its coordinates conceptually all play a role in determining what the sentence says in that situation. The coordinates of the circumstance of evaluation, on the other hand, capture each of the individual factors that are relevant for determining the truth or falsity of a sentence once what it says has been contextually determined. Once these two groups of parameters have been distinguished, it is possible to define two distinct but tightly related notions of meaning. The *content* of a sentence in context corresponds to the intuitive notion of what a sentence says on a given occasion of use. The *character* of a sentence, on the other hand, is the context-invariant content-determining rule that it is conventionally associated with and is what an agent must grasp if she is to be counted as competent in its use.

While this distinction is principally stated in terms of sentences, this is more or less just an accident of Kaplan's interest in developing a *logic* of demonstratives since any logic of this kind would need to be principally concerned with an analysis of the sentential notions of logical consequence and logical truth. The distinction seems to be conceptually more at home when analyzing lexical items. In particular, we now seem to have the intellectual resources needed to make sense of the puzzling features of English indexicals. The descriptive dimension of their meaning is due to their variable characters – in any given context, they may determine different contents. And this descriptive dimension appears to exhaust their meaning because the contents they determine are not variable with respect to the circumstance of evaluation.[1] Once context has fixed their values, they are fixed for good.

[1]In fact, Kaplan famously had the stronger intuition that their contents were not merely rigid in this sense, but that they were somehow particularly *direct* – something like the idea that the content of an indexical in context just is its referent, or that its referent determines its content (and not *vice-versa*).

Unsurprisingly, modeling the distinction between context and utterance requires more than merely saying that the conceptual roles they play are distinct and leaving it at that. Since they will inevitably individually function in more or less the same way from a technical perspective, faithfully representing the distinction requires that we enforce an important global architectural constraint on the design of our semantic theory. Everyone agrees that a semantic theory ultimately needs to be compositional in the rough sense that the meaning of a complex expression is functionally determined by the meanings of its parts and their grammatical arrangement. And assuming this constraint is enforced for contents, the fact that it will also hold for characters follows as a more-or-less trivial matter of course. But the implication does not hold in the other direction. So you can demand that the theory is *strongly* compositional, in that both contents and characters are assumed to satisfy some form of compositionality, or only *weakly* compositional in that compositionality is only enforced for characters.

Attaching any conceptual value to Kaplan's distinction requires that you endorse the stronger principle. The result is a multi-staged picture of interpretation that can be diagrammatically depicted as follows:



In the first stage, the characters of individual lexical items together with context determine the basic inputs to the compositional process, formally capturing the idea that context just serves a content-determining role. In the second stage, those lexical contents are compositionally assembled into the content of the overall sentence as used in that context. The result then determines a truth-value together with a circumstance of evaluation, formally capturing the idea that the circumstance of evaluation plays a truth-evaluative role. Something like this general picture presumably informs Kaplan's prohibition of what he calls *monsters* – non-trivial semantic operators that take characters for arguments. For a mon-

ster to make conceptual sense, semantic composition would need to occur before the contents of individual expressions have been resolved. But since composition explicitly occurs *after* context has already resolved the contents of individual lexical items, nothing remains for a character-operator to potentially operate on.

The machinery underlying Kaplan's theory was nothing new at the time. The observation that we had to evaluate the truth of a sentence with respect to multiple groups of parameters of roughly the same kind had been around since at least Kamp (1971), who pointed out that an adequate logical account of the word `now` seemed to require tracking time (at least) twice:

(48) `All those now alive will eventually be dead.`

Very roughly, evaluating an utterance of the sentence in (48) requires that the extension of the predicate `alive` be determined with respect to the time of that utterance and that the extension of the predicate `dead` be determined with respect to some time further in the future. Formally speaking, examples like this one were treated in so-called *two-dimensional* modal logics that evaluate the truth of a sentence in a model with respect to pairs of the same kind of parameter.[2] The absence of monsters in Kaplan's framework distinguishes it from these other more general ones. The advantage is that in enforcing a sharp conceptual distinction between the different roles that the coordinates of these pairs play, we are also giving an *explanation* for why a parameter of a certain kind (i.e., time, world, etc.) might need to occur multiple times in our formal semantic theory. One occurrence is needed to play the content-determining role, and one occurrence is needed to play the truth-evaluative role. In so doing, we avoid positing a mysterious notion of two-dimensionally parameterized truth. The content of a sentence on an occasion of use is the primary bearer of truth and falsity, and the truth of a content is only evaluated with respect to at most one occurrence of a parameter of any given kind.

Once Kaplanian contents are further identified with concepts of traditional interest in

[2]See e.g. Segerberg (1973) and Davies and Humberstone (1980)

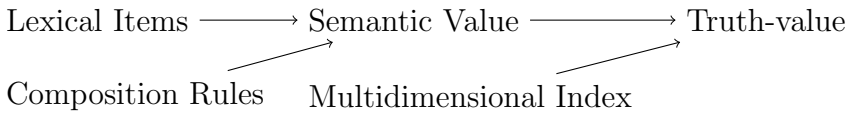
the philosophy of language, strong compositionality also becomes an extremely powerful methodological tool for investigating those concepts. For instance, if the content of a sentence in context is identified with the proposition that the sentence has been used to express, then we can argue for certain claims about propositions by considering the embedding behaviors of the sentences used to express them. Kaplan is normally regarded as having argued for *temporalism* about propositions in this way, in virtue of the presence of apparent tense-operators in English:

(49) a. `I am not speaking.`

b. `Sometimes, I am not speaking.`

It at least initially seems as though (49a) is a syntactic constituent of (49b), and so in any given context, strong compositionality requires that the content of (49b) is functionally determined in part by the content of (49a). And since (49a) can never be spoken truthfully, but (49b) can, it seems like the meaning of the word `sometimes` needs to be able to shift the time of evaluation away from the present. But in order for that to yield the correct results, the content of (49a) needs to vary with time. So, generally speaking, the truth of a proposition varies with time rather than being either eternally true or eternally false.

Unfortunately, this form of argument is also a double-edged sword. Lewis (1980) argued that if we *cannot* identify another role for Kaplanian contents to play outside of serving as an intermediate step in the interpretation of a sentence, they are theoretically superfluous. And the fact that they can be seemingly shown to vary with respect to so many different parameters in the way just illustrated rules out all of the plausible candidates. The resulting weakly compositional picture is one where any semantic interest in the context/circumstance distinction fades away:



According to this picture, semantics only traffics in *compositional semantic values*, which are generally sensitive to a vast array of different semantic parameters that all broadly serve the same truth-evaluative semantic role. Thus, unlike the Kaplanian picture, it puts a great deal of distance between the entities that ordinary philosophy of language traffics in and the entities that do most of the serious work in compositional semantic theory. In doing so, it also makes it extremely difficult, if not impossible, to directly argue for any interesting philosophical claim based on those kinds of considerations. It thereby removes a valuable and important form of argument from the philosopher of language's toolkit.

Therefore, it seems worthwhile to try and endorse some form of strong compositionality. The literature on responses to Lewis's form of the argument is vast, and I will not touch on that issue here.[3] Instead, I will be concerned with a recent generalization of Lewis's argument advanced by Rabern (2012, 2013) and Yli-Vakkuri (2013). These authors argue that the semantics of quantification already violates strong compositionality in a way that precludes anything like the general Kaplanian semantic architecture. While I think this argument can ultimately be resisted, careful attention to it will also yield surprising conclusions both about the contents of quantificational expressions and the nature of interpretation more generally.

5.2 Monstrous Binding

The arguments of Rabern (2012, 2013) and Yli-Vakkuri (2013) are built around a very basic observation about how quantifiers are normally handled in formal languages. For present purposes, we can consider a simplified semantics for the language of monadic first-order modal logic without individual constants. Models for this language are triples $\langle W, D, V\rangle$ consisting of a set of worlds W, a fixed constant domain of objects D, and a valuation V that assigns each predicate expression some function from W to $\wp(D)$. Against a fixed model

[3]See e.g. King (2003) and Glanzberg (2011) for the standard argument that the semantic value of a sentence in context can in fact be identified with the *eternalist* proposition that it expresses. Thus, it is an argument against both Lewis and Kaplan, who agreed that it would vary with respect to time.

$\mathcal{M}$, the standard definition of satisfaction then runs as follows:

$$
\begin{array}{lll}
w, g \vDash_{\mathcal{M}} Px & \text{iff} & g(x) \in V(P)(w) \\
w, g \vDash_{\mathcal{M}} \neg\phi & \text{iff} & w, g \nvDash_{\mathcal{M}} \phi \\
w, g \vDash_{\mathcal{M}} \phi \wedge \psi & \text{iff} & w, g \vDash_{\mathcal{M}} \phi \text{ and } w, g \vDash_{\mathcal{M}} \phi \\
w, g \vDash_{\mathcal{M}} \forall x\phi & \text{iff} & w, g' \vDash_{\mathcal{M}} \phi \text{ for every assignment } g' \sim_x g \\
w, g \vDash_{\mathcal{M}} \Box\phi & \text{iff} & w', g \vDash_{\mathcal{M}} \phi \text{ for every world } w' \in W
\end{array}
$$

Notice that the satisfaction relation is defined with respect to two different semantic parameters – a possible world w and a variable assignment g. At this point, we may ask which of Kaplan's two possible semantic roles each of them play. In order to answer this question, we first need to know what the correct notion of content should be for formulas of this language. Thankfully, this is reasonably straightforward. The content of a sentence of first-order modal logic is standardly assumed to be the *proposition* that it expresses, which is theoretically represented by some set of possible worlds (intuitively, the set of worlds where the proposition is true).

Against this background, contents are something whose truth is variable with respect to a world. That means that the world parameter must play an evaluative role, as it appears to merely supply the information relevant for assessing a content's truth at a particular point in logical space. Also, notice that it is impossible to directly assign any fixed content to an arbitrary formula of this language, as any free variables must first be assigned determinate values from the underlying domain before we can start calculating which proposition the formula expresses. So since the variable assignment provides this additional needed information, it seems right to assimilate it to the content-determining role instead. What we have so far said exactly follows Kaplan's own remarks on this matter – Kaplan (1989a) is quite explicit about the fact that the semantic role he assigns to the parameters of the context of utterance is directly modeled on the role that a variable assignment plays in supplying values to free variables. If right, then it seems best to model the character of a formula of this language as a function from variable assignments to propositions.

Now that we have a reasonably good handle on how the character/content distinction should apply in this case, we may ask whether or not the standard semantics for the language is strongly or weakly compositional.[4] Unfortunately, a quick examination of the clause for universal quantification suggests that weak compositionality is the best we can hope for. The problem is that strong compositionality would, in this case, require that the content of the expression $\forall x Fx$ with respect to g be the result of applying some operation O to the content of Fx with respect to g. But this is not what we, in fact, see. Instead, the correct account of universal quantification involves looking at the content of Fx with respect to a large family of many different assignments functionally related to g – the content of Fx with respect to g alone is in general insufficient for an adequate treatment. Hence, the quantifier phrase $\forall x$ appears to be monstrous since it needs to take an argument that has yet to be evaluated for its contextual value.

While quite a bit more could be said on this issue, we seem to have a good *prima facie* case for the claim that the standard approach to quantification in formal languages is incompatible with the key aspects of Kaplan's semantic picture that set it apart from its competitors. But should this by itself be any cause for concern? After all, Kaplan's framework is principally designed to explain why certain context-sensitive natural language expressions behave the way they do. And while we may productively use formal languages to study natural ones, the fact remains that they bear very little surface similarity to each other. In employing a formal language to model some natural language phenomenon, we must be prepared to accept that this mismatch might make a perfect transfer of concepts from one case to the other impossible.

The situation we presently find ourselves in seems to be a good case in point. Natural and artificial languages express quantification in two very different ways. For example,

[4]As we have already noted in Chapter 1, the answer to this question as things stand is *strictly speaking* that it is neither. Since a recursive definition of satisfaction does not directly assign a meaning to any expression (and is hence not a meaning assignment of any kind), it is trivially not a *compositional* meaning assignment. But this obstacle is a mere technicality, and we could easily convert this definition of satisfaction into a suitable meaning assignment using the technique employed in that chapter.

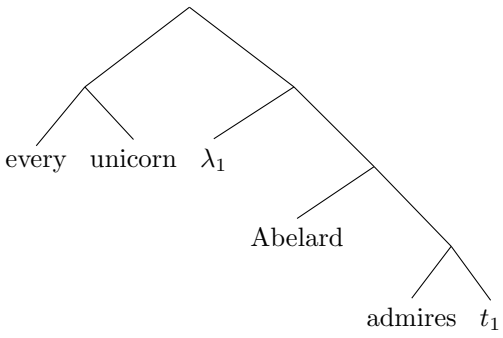
English employs a single quantifier phrase like `everything` that manifests in the surface position it directly controls, whereas formal languages of the kind we are presently considering require (at least) two grammatically disconnected units – a quantifier phrase $\forall x$ to mark scope, and a variable x to indicate the positions that quantifier governs. While splitting up quantification like this goes a long way towards increasing legibility and expressive agility, it also undeniably introduces technical complexities. In particular, we now need to design a semantic theory that allows a quantifier phrase to be semantically related to a distant and structurally unrelated variable it controls. And as we have already extensively discussed, variable assignments happen to be an extremely elegant and technically convenient way to solve this awkward problem. But importantly, they solve a problem that exists only because the designers of the language have chosen a particular idiom for the expression of generality. They are present in the semantics not because of *what* can be said, but because of *how* what can be said is said. Hence, it should be unsurprising that they do not fit neatly into Kaplan's framework. They may be a mere artifact of a decision in the design of an artificial language that was made out of convenience as much as anything else. They do not necessarily reflect anything of interest in the linguistic reality that Kaplan was interested in modeling. Of course, the above reasoning is convincing only if natural languages have no true structural analog to the variable-binding idiom. If there is such an analog, then it might turn out that we need to regard certain natural language expressions as sensitive to something like a contextually supplied assignment parameter to describe their semantic properties adequately. Providing a philosophically satisfying account of that parameter's conceptual role in our overall semantic theory would be just as pressing an issue as providing a similar account for any other. Thus, we may need to take the threat that assignment shifting operators pose to the Kaplanian interpretive architecture more seriously.

According to Rabern (2012, 2013), standard accounts of quantification in natural languages already give us sufficient reason for pause. As it turns out, a basic structural feature of quantification captured by the normal variable-binding idiom is likewise an issue that

needs careful handling when we turn our attention to natural languages. In particular, we still need an explanation of the double-life that natural language quantifier phrases seem to live. On the one hand, they obviously play an essential syntactic role in saturating one of a verb's open argument positions. Here, it is natural to think that they should likewise make some semantic contribution that registers that they directly control the argument position they have saturated. On the other hand, they also appear to need to take semantic effect somewhere *else* in the sentence, as they will sometimes need to take scope in a position that cannot always be immediately inferred from where they were first introduced in the syntactic derivation. As we have already noted, formal languages associate two different expressions to the two different semantic roles – the argument-saturating x and the scope-taking $\forall x$. Textbook treatments of natural language quantification appear to take a cue from this approach. More precisely, they assume that there is a level of syntactic description distinct from surface structure in which quantificational expressions may have been moved out of their surface positions:

(50) a. `Abelard admires every unicorn.`

b.



Very roughly, the quantifier phrase in (50a) is first introduced in the syntactic derivation in order to satisfy the requirement that the main verb of the sentence has a direct object. It is then moved out of that location by some mechanism, leaving behind a variable-like trace. This trace is then bound at the location that the quantifier takes scope by an accompanying abstraction operator, resulting in the structure depicted in (50b). Hence, something extremely close to the quantifier/variable dichotomy seems to be likewise posited by our best

semantic theories of natural language quantification, too. The relation between a trace and the abstraction operator that binds it seems to do the trick.

This way of making trouble for the Kaplanian minimally assumes that the movement-based approaches to this issue are empirically correct. Unfortunately, while this kind of theory is popular enough to enjoy default status, it is still far from a settled issue. There are quite a few alternative proposals actively being developed and investigated, many of which do not posit anything like covert bound traces. Moreover, it is an extremely tricky issue to distinguish these different approaches empirically, and so the scales are not likely to definitively tip in any one direction anytime soon.[5] But more importantly, it is still extremely unclear if there is an issue present, even if you do ultimately accept one of the movement-based approaches. We have a case for a potentially monstrous operator only if we can make sense of an expression with a *free* variable-like element as a fully interpretable independent unit of semantic analysis. Only then could we meaningfully assign this unit a content that we could point to as an essential ingredient of a potential exception to strong compositionality. While this is an extremely natural thing to say when you are considering the semantics of an artificial language that generates open formulae in the same breath as closed sentences, it seems like a very bad fit for the kind of syntactic theories that are assumed to generate structures like (50b). Given what we have said, nothing like a free variable is ever present as an independent unit in this kind of theory – binder and bindee are introduced in the structure simultaneously when a quantifier expression is moved out of position and the trace it leaves behind is bound by an abstraction operator. And since the derivational history of (50b) suggests an extremely tight syntactic connection between bound and binding elements, why should we endorse a semantic theory that nonetheless seems to treat them as distant and essentially unrelated?

Here is another way to think about this basic problem. For this argument to work, we need

[5]See Szabolcsi (2010) for a recent survey of a few different alternatives, as well as a discussion of how one might go about empirically distinguishing the different proposals from one another.

to assume that sentences like (50a) have an implicit context-sensitive indexical component despite surface appearances otherwise. Moreover, this component never *actually* functions as an paradigm indexical (e.g. a directly referential context-sensitive expression). It only exists to be operated on in the computation of the semantic value of a larger expression. In other words, it exists *only* to be the target of a monstrous operator. So in any given context, its real semantic contribution will not be the value it assumes in that context. But if that is the case, what reason do we have to treat them as context-sensitive expressions in the first place? While these considerations are not completely decisive against this particular strategy, they suggest that a different approach would be dialectically much more compelling. From the previous discussion, it is clear that the apparent problem stems from the existence of free-variable-like elements that a grammatically unrelated quantifier can then bind. We can, therefore, strictly improve this argument if the element in question is an explicit aspect of structure rather than a theoretical posit, which has uses that appear to be *genuinely* free.

Pronouns appear to exhibit all of the relevant features. Occurrences of pronouns have been traditionally divided into at least three different categories – deictic, bound, and anaphoric:

(51) a. `He is beautiful.`

b. `Every unicorn`^{1} `thinks that he`$_{1}$ `is beautiful.`

c. `Abelard`^{1} `thinks that he`$_{1}$ `is beautiful.`

In (51a), the pronoun functions more or less as a demonstrative and refers to some contextually selected salient individual. On the other hand, the pronoun in (51b) functions very much like bound variables do in predicate logic. So unlike (51a), it does not refer to a definite individual and inherits its meaning from its surrounding linguistic environment rather than the extra-linguistic context. The pronoun in (51c) is something like a mixed case – while it patterns with (51a) in that its content appears to be a definite individual, it patterns with (51b) in that it seems to inherit this content from its surrounding linguistic material rather

than the extra-linguistic context.

For our present purposes, we can note that bound and deictic readings together give rise to a problem. On standard Kaplanian assumptions, the content of (51a) in a context where (for example) Abelard is the most salient object should be the singular proposition that Abelard is beautiful. But it also seems that in this context, this proposition plays no obvious role in the compositional semantics of (51b) on the indicated reading where a hierarchically superior quantifier binds that pronoun. Therefore, since (51a) is a syntactic constituent of (51b), it looks as though something monstrous has occurred – the quantifier phrase has taken an argument in which the pronoun has not yet been evaluated for its contextual value. More explicitly, we take the observation that there is a context in which both (51a) and (51b) can be felicitously uttered to express their indicated readings to be uncontroversial. That is to say, there is a possible situation such that if a speaker were to sincerely and literally utter (51a), she would convey a piece of singular information, and if she were to sincerely and literally utter (51b), she would convey a piece of non-singular information instead. Since these bits of information are what would be conveyed by sincere and literal utterances of these sentences in this situation, we may plausibly assume that they are also their respective contents. But if we also grant that (51a) is a true syntactic constituent of (51b), then the following three principles are straightforwardly incompatible:

(**A**) The distinction between free and bound occurrences of English pronouns is not explained in terms of a lexical ambiguity

(**B**) The content of an expression is functionally determined by the contents of its parts

(**C**) In a given sentence, every lexical item's content is retrieved with respect to the same initial context of utterance

To see the problem, just note that according to (**A**), the general interpretive rule governing the occurrences of `he` in (51a) and (51b) would be the same. So, by (**C**), these occurrences

must also have the same content in the context at issue. But by (**B**), in order to express the piece of singular information conveyed by the possible utterance of (51a, that content must itself be singular. And so (again by (**B**)), it would be unsuited to express the piece of non-singular information conveyed by the possible utterance of (51b).

In its barest form, the argument against the Kaplanian framework that we have been discussing merely notes that the textbook treatment of pronouns avoids this problem by abandoning (**B**). But since (**B**) is just a restatement of strong compositionality, this means that the standard semantics of pronouns treats them as being sensitive to monstrous operators of some kind. But stating the argument in this way also makes it painfully clear that this conclusion is not inevitable. In particular, one may choose to abandon either (**A**) or (**C**), instead. But while abandoning (**A**) has recently proven to be a somewhat popular strategy, to my knowledge, no one has seriously pursued abandoning (**C**). So in what follows, we will instead sketch how one might experimentally develop a response of this kind, using the theory of discourse referents presented in the previous chapter as a foundation.

5.3 The Character of Pronouns

As we just noted, occurrences of pronouns are traditionally placed into one of three different categories based on their semantic behavior: deictic, bound, and anaphoric. Most analyses of anaphoric occurrences then try to assimilate them into one of the other categories, leaving just two genuinely distinct types.[6] It is then standardly assumed that the distinction between deictic and bound occurrences should be modeled on the analogous distinction between free and bound occurrences of variables in most familiar formulations of predicate logic. But as we saw in Chapter 1, there are essentially two quite different views that one could take with respect to that formal distinction, and so two quite different corresponding views of the

[6]For classic discussions, see, e.g., Lasnik (1976) for a view that groups anaphoric occurrences with deictic ones and Evans (1977) for a view that groups (syntactically local) anaphoric occurrences with bound ones.

pronoun occurrences it is used to model.

According to the Tarskian perspective, free and bound variables are given a univocal semantic treatment, and so on that model, all pronoun occurrences are governed by the same general kind of Tarskian rule. If that rule is further understood as giving the character of the pronoun occurrences in question, the result is the standard (and monstrous) analysis. On the other hand, the Fregean perspective does not officially recognize that free occurrences of variables are legitimate. After all, the Fregean grammar for first-order predicate logic does not generate open formulas. So adapting the Fregean view to apply to a more general class of formulas requires positing a quite radical ambiguity. In particular, bound occurrences may be treated as they normally are in the Fregean language. That is to say, they are seen as pieces of uninterpreted punctuation, which serve the grammatical role of indicating lateral quantifier-control structure in much the same way that parentheses serve the grammatical role of indicating hierarchical constituent structure. Free variables, on the other hand, will need to syntactically function as singular terms and serve to (directly) designate the specific object assigned as its value.

This Fregean-inspired vision of the distinction between the free and bound occurrences of variables directly informs a corresponding vision of the distinction between deictic and bound occurrences of pronouns. According to this perspective, deictic occurrences are governed by an interpretive rule that allows them to (directly) refer to some individual supplied by context and so are best modeled by assigning them non-trivial characters. Bound occurrences, on the other hand, are governed by a completely different kind of interpretive rule that treats them as mere placeholders in the construction of complex predicates. Importantly, pronouns are only allowed to serve this role in a very specific kind of grammatical configuration. In particular, there must be some antecedent linguistic material that coerces it into making this special contribution. Normally, that would be a structurally superior quantifier phrase, though, in principle, other kinds of configurations may also license this behavior. It is easy to think that occurrences of pronouns that serve this role are somehow contentless. Like the

Fregean interpretation of bound occurrences of variables, they function merely as a way of broadcasting that the argument positions they occupy are somehow structurally governed by their licensing antecedents rather than as serving as independently interpreted units in their own right.[7] This view of pronouns, therefore, avoids monstrosity at the cost of positing an exotic form of ambiguity that only pronouns seem to participate in.

In order to avoid positing either monstrosity or a special form of ambiguity, my stance towards the traditional tripartite classification of pronoun occurrences will be slightly different. To reiterate, the views just outlined assume that pronouns really have just two core semantic behaviors, exhibited by paradigmatic deictic and bound occurrences, respectively. Anaphoric occurrences are regarded as special cases of either deictic or bound occurrences, depending on other background theoretical preferences. I will invert this typical explanatory strategy to find a univocal and non-monstrous interpretive rule governing all pronoun occurrences. In other words, I will take the semantic behavior of anaphoric occurrences as the paradigm to which the other two categories should be assimilated, rather than the other way around.[8]

---

[7]The remarks here have been largely informal, and this picture can be formally developed in several different ways. I have suggested that proponents of this view believe that bound pronouns should receive something like a syncategorematic treatment, though there are other formal possibilities in a similar spirit. For example, Salmon (2006) and Glanzberg and King (2020) can be understood as developing very similar proposals that allow structurally superior quantifiers to coerce pronouns into making a different semantic contribution than the one that it would have made in some other environments. But these proposals still treat these occurrences categorematically, in the sense that they are still assigned a definite and independent semantic value (though one which in practice may only be deployed in a specific kind of setting).

[8]In this respect, the mechanics of my proposal are extremely similar to what is proposed in Stojnić et al. (2017). These authors present a theory that (on my terminology) treats deictic occurrences directly on the model of anaphoric ones (rather than the other way around, which is much more typical). But our interests are otherwise quite different. They are mainly concerned with providing the schematic details of how the contextual salience of a discourse referent evolves in context, and I have nothing of real substance to say about that. And they have relatively little to say about what the contents of pronoun occurrences in context, in fact, are. Their discussion primarily focuses on bread-and-butter cases of cross-sentential anaphora with singular (including indefinite) antecedents. And this at least suggests they are thinking of the contents of these occurrences as particular entities, along the same lines suggested in Chapter 3. On the other hand, I am explicitly concerned with cases where the antecedent of a pronoun is a possible plural quantificational expression and where that antecedent may also be syntactically local. In these cases, we have seen that the content of a pronoun is best identified with a variable-meaning. I am thus explicitly showing how this style of account can be generalized to an even broader class of pronoun occurrences, so long as we accept the

Thankfully, we have already extensively sketched a theory of how to interpret anaphoric occurrences of pronouns. According to the view presented in the previous two chapters, an occurrence of this kind has its content fixed by one of the discourse referents introduced by its antecedent, where discourse referents are then understood as ways of presenting the subjects of conversation. This basic idea can be easily incorporated into a standard Kaplanian framework with the following interpretive rule:

> (**Anaphoric Character**) The content of a pronoun occurrence π with antecedent α in context c is the value of the discourse referent δ associated to α by the context c in the world of c.

This principle specifies the character that governs anaphoric occurrences and is theoretically neutral in the sense that it does not place any real constraints on what realizes the discourse referent role (aside from the fact that they have values in a world). So the range of pronoun occurrences this principle can account for largely depends on how those further details are fleshed out. And it seems like our particular theory of discourse referents has all the resources needed to subsume deictic and bound occurrences of pronouns, too.

The basic idea behind our proposal was the observation that a speaker is normally expected to have certain things in mind whenever they make a warranted assertion. This was codified as the (**QP Subjects**) principle, which stated that we could, in particular, ordinarily assume that a speaker has (at least) the domain of any quantifier she used together with suitable witnesses to the truth of their statement ready at hand. The latter component was the primary focus of much of the discussion that followed. In particular, we noted that a witness to the truth of a claim involving a polyadic quantifier normally takes a relational form. And an n-ary relational witness could equivalently be modeled as a sequence of n distinct variable-meanings that jointly represented the witness against the background of a suitably

---

proposal about their contents argued for in Chapter 4. We thus share broadly the same outlook, though we have chosen to more systematically develop different parts of the picture first.

general state-space. This way of fragmenting a relational witness allowed us to uniquely associate variable-meanings to each of the noun phrases that contributed to the expression of a complex polyadic quantifier. Those phrases could then be seen as the antecedents of latter anaphorically linked pronouns that took those associated variable-meanings as their contents. By putting that all together, we arrived at the idea that discourse referents were variable concepts defined in terms of a speaker's communicative intentions. As a simple example, it could account for cases of plural anaphora like the following:

(52) a. `Most of my students`[1] `failed the final exam. They`$_1$ `didn't study.`

b. [MOST $\mathtt{v}_1$: STUDENT $\mathtt{v}_1$] FAIL $\mathtt{v}_1 \wedge \neg$STUDY $\mathtt{v}_2$

The fact that `Most of my students` is co-indexed with the subsequent pronoun `they` indicates that this pronoun's content was fixed by a discourse referent associated with that antecedent expression. On its most natural reading, this content is a variable-meaning v whose value range is restricted to all and only the witnessing students that a speaker of (52a) had in mind. If the variable $\mathtt{v}_2$ is stipulated to be interpreted by v, the truth-conditions of (52a) are then straightforwardly representable by the open formula (52b), so long as that formula is given its ordinary (universally closed) interpretation.

On this kind of picture, it is more or less trivial to also account for deictic occurrences of pronouns, so long as we suppose that they always co-occur with some kind of content-fixing demonstration. Those demonstrations can then be seen as the antecedents of the pronouns in question, introducing discourse referents whose contextual values are once again variable-meanings with appropriately restricted ranges.[9] And the expectation of an accompaniment of this kind is perfectly reasonable. After all, coordination on the correct proposition in cases like this would be straightforwardly impossible in the absence of any suitable linguistic cue. For instance:

(53) a. `They didn't study.`

[9]See Stojnić et al. (2017) for a much more extensive and sophisticated development of this idea.

b. $\neg\textsc{study}(\mathtt{v}_1)$

If a speaker utters (53a) completely out of the blue, without suitable antecedent linguistic context or accompanying content-fixing demonstration, an interpreter will be completely unable to determine the subject of the speaker's assertion. But, if the speaker additionally points to a stack of disappointing exams concurrently with their utterance, a competent interpreter will determine that the speaker intends to talk about the students who took those exams more or less automatically. According to the present account, that demonstration can be seen as introducing a discourse referent whose contextual value is some variable v ranging over all and only those students. The content of the speaker's utterance in context is then once again straightforwardly representable as the (open) formula in (53b).

Accommodating bound occurrences, on the other hand, is much less straightforward. But it is possible, so long as we see the semantics of natural language quantification as itself involving variable-binding of some kind. With this caveat in place, a bound reading of a pronoun is predicted by our account to arise whenever that pronoun's antecedent is structurally superior to it in the same propositionally-valued construction. To see why this is the case, note that our account states that a discourse referent may track either a domain of quantification or a witness. Our focus on sentence-to-sentence discourse-level facts up to this point has made witness-tracking discourse referents much more pertinent. In the situation that we are presently considering, however, domain-tracking discourse referents look like they are the only coherent possibility. Witnesses help characterize one important aspect of a speaker's evidential support for an asserted proposition. They are the things that speakers can point to if called upon to justify themselves further and ideally suffice to establish the truth of their claims. So it follows that witnesses of this kind are only coherent against the background of conversationally salient propositions needing evidential support. But that means that witness-tracking discourse referents are only definable against the background of clausal antecedents that supply those propositions. Domain-tracking discourse referents, on the other hand, are not subject to this conceptual requirement – in fact, there is no real

formal obstacle to thinking that they are just what is immediately supplied by an explicitly restricting nominal. So in the kind of configuration that we are currently considering, it follows that a domain-tracking discourse referent is the only reasonable option since there is no plausible clausal antecedent for the pronoun occurrence in question. The result is something like the following:

(54) a. `Most of my students`^{1} `forgot their`$_{1}$ `blue books.`

b. [MOST $\mathtt{v}_1$: STUDENT $\mathtt{v}_1$] FORGOT $\mathtt{v}_1$ BLUEBOOKOF $\mathtt{v}_1$

When a speaker utters (54a), they provide an explicit restricting nominal phrase `my students`, which directly corresponds to the domain that the quantifier `most` ranges over. We can think of that nominal phrase as introducing a domain-tracking discourse referent whose value in the context of utterance is a variable-meaning v restricted to range over all and only the speaker's students, and which is controlled by a variable-binding operator introduced by `most`. This domain-tracking discourse referent is the only reasonable option for interpreting the anaphorically linked pronoun `their`. So this pronoun occurrence winds up reusing the variable-meaning v, resulting in a reading where the out-scoping variable-binding operator controls it. Taking the variable `v` to be interpreted by v, the expressed truth conditions are therefore representable as something like (54b).

I hope that the preceding brief discussion suffices to minimally establish the *prima facie* viability of (**Anaphoric Character**) as an informal statement of the univocal interpretive principle that governs all pronoun occurrences, given our background understanding of discourse entities. But we have also glossed over a difficult technical snag. In particular, we are now assuming that context encodes the information about a pronoun occurrence's suitable antecedents. But we also distinguished between novel anaphoric possibilities that are immediately made available and ones that are only made available against the background of a suitable clausal antecedent. In particular, our explanation of bound readings at least implicitly involves the idea that a domain-tracking discourse referent is made available in

context as soon as a restricting nominal is encountered. But formally accounting for this behavior may involve monstrous operators.

5.4 Character and Content in a Dynamic Context

To make the problem that we are now faced with especially vivid, we begin by considering the following particularly simple example:

(55) a. `Bill went shopping.`

b. `He bought cigarettes.`

A discourse-initial utterance of the sequence in (55) will ordinarily express the propositions that Bill went shopping and that Bill bought cigarettes. In order to formally capture this fact, a semantic theory needs to always deliver the reading where the occurrence of `Bill` is anaphorically linked to the pronoun `he` when the sequence is evaluated against a *null context* (i.e., a context that is assumed to contain no prior information about the discourse salience of any given entity). But designing a formal theory capable of delivering this result while also holding on to other important theoretical desiderata is harder than it sounds. Any such proposal will involve the specification of a *discourse sequencing* operator, which will explain how to compute the meaning of two concatenated utterances. From a purely truth-conditional perspective, this operator is clearly some kind of conjunction. But equally clearly, the contribution of a conjunction operator to the computation of a discourse's truth-conditions cannot exhaust this operator's meaning. Conjunction is always commutative, while the interpretation of discourse is generally not – uttering (55a) before (55b) normally expresses something different than uttering (55b) before (55a). Of course, this behavior is easily explained once we realize that utterances are literally context-changing in a certain minimal sense. Once a given entity has been mentioned, its salience to an interpreter is automatically promoted, thereby making it much easier to refer to it again in later utterances. So utterances of (55a) tend to increase Bill's salience. That means that he is the default most

reasonable candidate to serve as the content of the pronoun occurrence in (55b), whenever (55b) is uttered immediately *after* (55a). But if (55b) is uttered *before* (55a), then he will not be favored as an interpretive default in this way. The following interpretative rule straightforwardly captures this extremely plausible story:

> (;) The content of $\phi;\psi$ in context c is the conjunction of the content of ϕ in c and ψ in c', where c' is exactly like c except that all of the entities mentioned in ϕ have been promoted in salience.

But now we can easily see the problem – this rule obviously violates strong compositionality. In order to determine the content of $\phi;\psi$ in c, one must determine the content of ψ in a shifted context c'. In other words, the discourse sequencing operator appears to be a monster.

At this point, you might be tempted to get out of this problem by denying that these general facts about the interpretation of sequential discourse need to be explained by semantics, relegating them to the proverbial wastebasket of pragmatics instead. On this way of thinking, there would be no need to give a formal semantic analysis of the discourse sequencing operator, and so no need to admit monsters into the theory. But this way of thinking is a bit too short-sighted. In fact, the general problem is a little more pervasive than the above discussion suggests. For instance:

(56) `Bill thought that he bought cigarettes.`

A discourse-initial utterance of (56) likewise always seems to result in a reading where the occurrences of `Bill` and `he` are anaphorically linked. And the explanation of this phenomenon is not appreciably different – the fact that Bill is mentioned earlier in the utterance promotes his salience, making him the only reasonable candidate to serve as the content of the pronoun in the absence of any other discourse information. But notice that if we were to design a theory that formally incorporated something like this reasoning, we would not just be strictly localizing the source of monstrosity to the semantics of a sentence-sequencing operator. Instead, our story seems to require a context shift as soon as the proper name `Bill`

is encountered. And this kind of intrasentential shift is not easily handled by archetypal theories of pragmatic inference and context update, *à la* Grice (1975) and Stalnaker (1978). But semantically delivering this result is also quite difficult, as it would seem to require that we treat even *proper names* as monstrous operators of sorts in an otherwise normal Kaplanian setup.

To get a better handle on this problem, it will be useful to consider a typical textbook discussion of anaphora like the one in Büring (2005). According to the elementary theory presented there, occurrences of proper names and pronouns are assumed to carry numerical indices, and context is assumed to supply an assignment function mapping these indices to entities from the underlying domain. Indexed proper names and pronouns are then supplied lexical entries like the following:

$$\begin{aligned} [\![\texttt{Bill}_7]\!]^g &= g(7) \quad \text{if } g(7) \text{ is Bill, and undefined otherwise} \\ [\![\texttt{he}_5]\!]^g &= g(5) \quad \text{if } g(5) \text{ is male, and undefined otherwise} \end{aligned}$$

This basic theory treats proper names and pronouns as traditional variables of sorts, which carry with them certain contextual presuppositions (identity presuppositions in the case of names and gender presuppositions in the case of pronouns). In this framework, the phenomenon that we are now interested in explaining is strictly speaking a kind of *parsing* issue, rather than a semantic one. Other things being equal, why do we tend to assign the same indices to the occurrences of the name and pronoun when (55) and (56) are encountered discourse-initially? In other words, it is really just a presemantic issue of contextual disambiguation and so may safely be ignored when giving the semantic machinery for interpreting anaphora.

On this way of thinking about things, formal contexts (assignment functions) merely record all of the referential possibilities that proper names and pronouns have. Without further modification, they do not provide any information that might be helpful to an agent that is actively trying to interpret these kinds of expressions as they are encountered (unin-

dexed) in the wild beyond the bare definedness criteria. But clearly, that information must be present in the conversational environment in some capacity. Even the staunchest advocate of this approach will have to grant that much. After all, how else could an intelligent parser normally arrive at contextually reasonable patterns of indexation? The difference between the apparently monstrous proposal that we were earlier considering and more mainline proposals, therefore, largely reduces to a difference in how this information is formally handled. Is it explicitly represented as a predictable and well-behaved aspect of context that the meanings of expressions are partially sensitive to, or is it instead seen as serving some non-semantic role in situated parsing? But despite appearances to the contrary, the difference between these two outlooks does not actually amount to all that much.

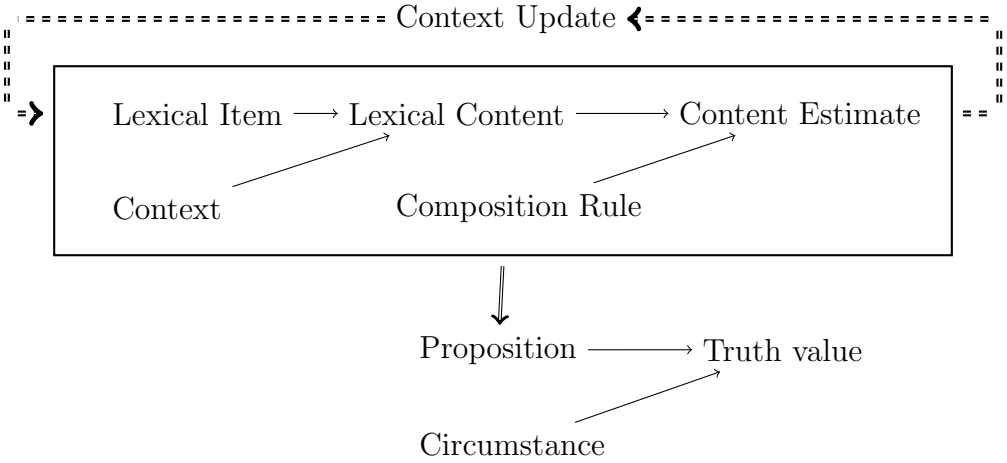
To see this last point, remember that situated interpretation is ordinarily analyzed as involving many different discrete processes. Some familiar examples of these are *parsing*, *compositional evaluation*, and *contextual enrichment*. Moreover, these processes are often implicitly ordered by the information they presuppose is readily available. Parsing, for instance, normally presupposes that some perceptible stimulus has been suitably processed and segmented into a string of lexical items, while compositional evaluation presupposes that the same perceptible stimulus has already been assigned a hierarchical structure. Taken literally, the result is a picture of interpretation that unfolds in sequentially executed stages. In the initial stages, a sentence is encountered in the environment and syntactically analyzed. In the following stages, the interpreter assigns the sentence a literal meaning that is functionally derived in part from the structure that it was earlier assigned. And in the final stages, the message that the speaker intended to convey is calculated from the literal meaning of the sentence they uttered together with other prevailing contextual factors.

In this kind of theoretical architecture, the place where information pertinent to anaphora resolution is first deployed may seem conceptually significant – the earlier in the process this happens, the more (quasi-)syntactic the phenomenon appears to be. But while this kind of picture may be extremely useful as a theoretical idealization, it is also profoundly implau-

sible as a psychologically realistic model of human language processing. As we mentioned in chapter 1, interpretation is now, by and large, agreed to be *incremental*. So instead of happening in sequential order, interpretive processes are thought to happen concurrently, with the flow of information between them often being bi-directional. If parsing and compositional evaluation are processes that occur in tandem, then the difference that we have been considering up to this point is completely insubstantial. When the pronoun is first encountered in (56), for example, an agent employing an incremental interpretation strategy will need to immediately generate a hypothesis about the index that it bears so that a plausible meaning estimate of the text so far consumed can continue to be maintained. And clearly, this index will be chosen so that the occurrence of the pronoun has the most plausible content in context. Thus, a pronoun occurrence's index and content are always computed in the same breath, using the very same contextual information. Whether or not this information is *officially* used to compute an index from which a content is derived or vice versa is a distinction without difference – these two proposals are functionally indistinguishable. Moreover, it seems that the largely semantic perspective on this issue has won out, since it is hard to see the relevant information as doing anything *other* than playing a primarily content-determining role. And this means that a consistent modeling strategy will require us to assign a univocal character to pronouns that directly uses this information. The standard use of indices is just a way of hiding the problem rather than dealing with it head-on. But this also apparently just leaves us in the awkward position that we started with. In particular, it looks like we will need to posit monstrous operators of some kind since that information undeniably mutates over the course of the evaluation of a single sentence in reliable and systematic ways.

Or do we? Recall that strong compositionality is essentially the idea that a word's dependence on context needs to be resolved before applying a compositional rule. But the ordinary formulation of this principle also seems to presuppose something like a classical (i.e., non-incremental) interpretive architecture. In that setting, strong compositionality

really amounts to thinking that the process of calculating a sentence's content in context happens in a single multi-phased episode. In the first phase, a sentence is encountered in the environment and syntactically analyzed. In the next phase, the interpreter resolves any possible context-sensitivity and retrieves the content of each individual word. And in the final phase, composition rules are applied to these contents to derive the expressed proposition. In an incremental setting, on the other hand, strong compositionality is just the requirement that contents are computed *on demand*. In other words, when a word is being processed, its sensitivity to context must be resolved immediately (i.e., before the next word is processed). Therefore, incrementally interpreting a segment of text in a strongly compositional way involves many small cycles with essentially the same multi-phased structure. First, an occurrence of a lexical item is encountered in the environment. Then, its structural position is inferred, and its content is retrieved in context. And finally, some compositional rules are applied to functionally integrate that content into the running estimate of the content of the text consumed so far. The difference is that an incremental architecture makes room for context shifts to occur *between* each of these individual cycles. The result is the following modest revision of Kaplan's picture:



Notice that each cycle composition only occurs after a content has been computed. So we can avoid positing monsters in a theory of interpretation that systematically exploits the

gaps between these cycles in the right way.[10]

We conclude by briefly returning to the puzzle presented in §2:

(51) a. `He is beautiful.`

b. `Every unicorn`^{1} `thinks that he`$_1$ `is beautiful.`

The above pair looked like it might have been a counterexample to strong compositionality. But when we tried to flesh the problem out in greater detail, we saw that we could only draw the weaker conclusion that the following three principles were incompatible:

(**A**) The distinction between free and bound occurrences of English pronouns is not explained in terms of a lexical ambiguity

(**B**) The content of an expression is functionally determined by the contents of its parts

(**C**) In a given sentence, every lexical item's content is retrieved with respect to the same initial context of utterance

In §3, we presented a solid reason for thinking that we should hold on to (**A**). In particular, given our theory of discourse referents, it seemed possible to explain all pronoun occurrences as being governed by the single interpretive principle (**Anaphoric Character**). What we have now just done is show how it is possible to hold on to (**B**), so long as we recognize that context shifts reflecting the introduction of novel discourse referents can occur in reliable and predictable ways *after* each lexical item's content has been retrieved and integrated into a running meaning estimate. So, in essence, ambiguity and monstrosity can both be plausibly avoided at the cost of abandoning (**C**). More specifically, according to our proposal, pronoun occurrences will normally be evaluated against the background of a context that is informationally richer than the one that its antecedent was evaluated against. And

[10] A sketch of a proposal along these lines is given in Appendix C.

bound occurrences, in particular, will involve a kind of intrasentential context shift where a restricting nominal first introduces a new domain-tracking discourse referent, which is then resumed by a linked pronoun later down the line.

APPENDIX A

Incremental Interpretation of Predicate Logic

In Chapter 1, I argued that the Fregean approach to the semantics of predicate logic is incompatible with a simple left-to-right, symbol-by-symbol interpretation strategy. I also suggested that the Tarskian approach is straightforwardly compatible with an incremental interpretation algorithm. This appendix is dedicated to demonstrating this fact more explicitly. An incremental parser is just a device for recognizing some set of strings that works in a strictly left-to-right fashion. For our current purposes, we can think of an incremental interpretation procedure as an incremental parser that has been modified to assign an accurate meaning-estimate to the portion of the text it has so far consumed at any intermediate stage of a recognition process. Parsers are often usefully thought of as having three abstract components: a grammar, an algorithm, and (optionally) an oracle for resolving any implicit non-determinism in the algorithm. Since my claim is that the Tarskian theory is directly compatible with an incremental interpretation strategy, the first of these components is completely settled – the grammar that the parser needs to use is the same CFG presented in that chapter. And since it is well known that this grammar, at least, exhibits no structural ambiguity, it is easy to choose our algorithm to be entirely deterministic, eliminating any need for the third component.

Since it would also make my claim somewhat weaker if we had to use an algorithm that was fine-tuned to the specific idiosyncrasies of the grammar in question, we will instead use a simplification of the ready-made and well-studied left-corner parsing technique.[1] This

[1]See Grune and Jacobs (2008) §10.1.1 for details on the more general form a left-corner parser takes.

algorithm combines elements of top-down prediction and bottom-up matching, which allow it to easily work in a strictly left-to-right fashion. In order to formally state this algorithm, we need to specify a set of initial configurations that all computations begin in, a target configuration that successful computations terminate in, and rules which freely and non-deterministically apply to change one configuration into another.[2] A configuration can be abstractly thought of as a pair representing the memory state of a recognizing automaton and the freely available information in the environment. In this case, a memory state is any sequence of terminal and non-terminal symbols of the grammar (now thought of as the memory stack of a recognizing automaton). And an environmental state is any sequence of terminal symbols, which initially represents the string to be accepted or rejected. Here is the complete specification of the algorithm along these lines:

Left-corner Parsing Algorithm

Initial Configuration: $\boxed{S \mid w_1 \ldots w_n}$ **Target Configuration**: $\boxed{\mid}$

Match: $\boxed{\alpha\ X \mid w\ \beta} \Rightarrow \boxed{\alpha \mid \beta}$ where (1) $X \longmapsto w$ is a rule of the grammar, or (2) $X = w$

Predict & Connect: $\boxed{\alpha\ X \mid\ w\ \beta} \Rightarrow \boxed{\alpha\ Y_n \ldots Y_2 \mid \beta}$ where $X \longmapsto Y_1 \ldots Y_n$ is a rule of the grammar, and either (1) $w = Y_1$ or (2) $Y_1 \longmapsto w$ is a rule of the grammar.

Notice that each rule consumes the leftmost symbol environment, with no mechanism in place for forward-looking or backtracking. Hence, this algorithm recognizes strings in a strictly incremental fashion.

With the basic parser in place, we must now modify it to handle abstract semantic interpretation. The basic idea is extremely straightforward. Each rule in the grammar gets associated with a semantic object – lexical look-ups for lexical rules and simple meaning-recipes corresponding to the generic meaning type of the identified constituent structure for

[2]Again, as it turns out this potential for non-determinism is not realized in the specific case we are interested in.

structural rules. Here is the modified grammar:[3]

Interpreted Grammar for Incremental Interpretation

Structural Rules:

$[\![id_{\tau(Pred_n)}]\!] : S \longmapsto Pred_n \overbrace{T \ldots T}^{n\text{-times}}$

$[\![\neg]\!] : S \longmapsto \neg S$

$[\![\wedge]\!] : S \longmapsto (S \wedge S)$

$[\![\exists]\!] : S \longmapsto \exists\ Var\ S$

Lexical Rules:

$[\![c]\!] : T \longmapsto c\ (c \in \mathcal{C})$

$[\![v]\!] : T \longmapsto v\ (v \in \mathcal{V})$

$[\![v]\!] : Var \longmapsto v\ (v \in \mathcal{V})$

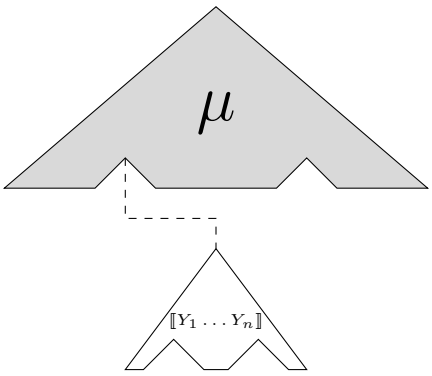
$[\![P]\!] : Pred_n \longmapsto P\ (P \in \mathcal{P} \text{ and } \rho(P) = n)$

Then we need to specify what we need to do with them when a step in the parsing procedure invokes the associated rule. But this is extremely straightforward since we only have two possible actions, which can each be given a mostly uniform interpretation. **Match** will generally tell us to perform function application, and **Predict & Connect** will generally tell us to perform a generalized form of function composition.

More precisely, our task is to associate an estimate μ of the meaning of the string thus far processed to each stage of the recognition process. This just requires that we fix an initial estimate μ_0 and show how to construct μ_{n+1} for any given μ_n and possible next step of the algorithm. The first task is easy. Since none of the string has been read in the initial configuration of the recognition task, we can safely choose some null semantic object. Thus, we set $\mu_0 = id_{s \to t}$, the identity function on the type of open formulas. **Match** steps are also fairly straightforward. Given the way our grammar is currently set up, every time **Match** is

[3]Notice, again, how the modified grammar more-or-less directly implements the Tarskian theory. Each rule is paired with either a trivial entry (for atomic sentences) or a meaning-object the standard version of the theory already made use of.

invoked, the matched symbol will either be an uninterpreted (or already interpreted) piece of syntax (cases where $X = w$) or the current meaning estimate μ_n will be a function whose argument type matches the lexical item that was just identified. In the first case, we can safely let $\mu_n = \mu_{n+1}$. In the second case, we have identified a lexical rule $X \to w$ and so just feed the meaning of that lexical item $[\![w]\!]$ to μ_n (i.e. $\mu_{n+1} = \mu_n([\![w]\!])$). **Predict & Connect** steps all correspond to an identification of some sentential constituent. Given the way our grammar is currently set up, if **Predict & Connect** is ever invoked, our current estimate of the meaning μ_n will be a partial specification of the meaning of the identified constituent's surrounding sentential context. Abstractly, you can think of this object as potentially containing many sentence-type holes, the leftmost of which is signaled out as the place where we must lower the newly identified constituent. But since we need to lower the meaning into its proper position immediately after it is identified (but before it has been fully processed), function application is somewhat inappropriate. Thankfully function composition (suitably generalized to arbitrary composable functions) does exactly the job we need.[4] Visually, this may be represented as follows:



The one hiccup is when clause (2) is invoked, which tacitly involves a match-step in addition

[4]More precisely, let $f : \alpha_1 \to (\ldots \to (\alpha_n \to \beta_1))$ and $g : \beta_1 \to (\ldots \to (\beta_m \to \gamma))$ be composable functions. Define their generalized composition $g \circ f : \alpha_1 \to (\ldots \to (\alpha_n \to (\beta_2 \to \ldots \to (\beta_m \to \gamma))))$ by letting $g \circ f = \lambda x_1 \ldots \lambda x_n \lambda y_2 \ldots \lambda y_m . g(f(x_1 \ldots x_n), y_2, \ldots, y_m)$. Technically, this defines a family of composition operators which already has a well established role in incremental semantic theories. In fact, they are just a very minor generalization of the semantic interpretations of the normal family of $\mathbf{B}^n$ combinators from Combinatory Categorial Grammar, which takes incrementality as one of its main selling points. See Steedman (2000) for more details.

to the standard connection step. In that case, we instead invoke the lexical item associated with the appealed-to lexical rule. Formally, we need to extend our parsing algorithm with an additional component that keeps track of the current meaning estimate and how to modify it at each stage. Given the above discussion, that is easy enough to do. Here is the specification of the semantically modified grammar:

Left-corner Parsing with Semantic Field

Initial Configuration: $\boxed{S \mid w_1 \ldots w_n}\ \boxed{id_{s \to t}}$

Target Configuration: $\boxed{\mid}\ \boxed{[\![w_1 \ldots w_n]\!]}$

Match (1): $\boxed{\alpha\ w \mid w\ \beta}\ \boxed{\mu} \Rightarrow \boxed{\alpha \mid \beta}\ \boxed{\mu}$

Match (2): $\boxed{\alpha\ X \mid w\ \beta}\ \boxed{\mu} \Rightarrow \boxed{\alpha \mid \beta}\ \boxed{\mu([\![w]\!])}$

where $[\![w]\!] : X \longmapsto w$ is a lexical rule of the grammar.

Predict & Connect (1): $\boxed{\alpha\ X \mid\ Y_1\ \beta}\ \boxed{\mu} \Rightarrow \boxed{\alpha\ Y_n \ldots Y_2 \mid \beta}\ \boxed{\mu \circ [\![Y_1 \ldots Y_n]\!]}$

where $[\![Y_1 \ldots Y_n]\!] : X \longmapsto Y_1 \ldots Y_n$ is a structural rule of the grammar

Predict & Connect (2): $\boxed{\alpha\ X \mid\ w\ \beta}\ \boxed{\mu} \Rightarrow \boxed{\alpha\ Y_n \ldots Y_2 \mid \beta}\ \boxed{\mu \circ [\![w]\!]}$

where $[\![w]\!] : Y_1 \longmapsto w$ is a lexical rule of the grammar and $[\![Y_1 \ldots Y_n]\!] : X \longmapsto Y_1 \ldots Y_n$ is a structural rule of the grammar

A complete trace of a single run of the algorithm is given on the next page for illustration:

	Current Configuration	Interpretation	Justification
1.	$\boxed{S \mid \neg\exists x(Rxa \wedge Px)}$	$id_{s \to t}$	Initial Configuration
2.	$\boxed{S \mid \exists x(Rxa \wedge Px)}$	$[\![\neg]\!]$	Predict & Connect (1)
3.	$\boxed{S\ Var \mid x(Rxa \wedge Px)}$	$\lambda v.\lambda\phi.[\![\neg]\!]([\![\exists]\!](v, \phi))$	Predict & Connect (1)
4.	$\boxed{S \mid (Rxa \wedge Px)}$	$\lambda\phi.[\![\neg]\!]([\![\exists]\!]([\![x]\!], \phi))$	Match (2)
5.	$\boxed{)\ S \wedge S \mid Rxa \wedge Px)}$	$\lambda\phi.\lambda\psi.[\![\neg]\!]([\![\exists]\!]([\![x]\!], [\![\wedge]\!](\phi, \psi)))$	Predict & Connect (1)
6.	$\boxed{)\ S \wedge T\ T \mid xa \wedge Px)}$	$\lambda u.\lambda v.\lambda\psi.[\![\neg]\!]([\![\exists]\!]([\![x]\!], [\![\wedge]\!]([\![R]\!](u, v), \psi)))$	Predict & Connect (2)
7.	$\boxed{)\ S \wedge T \mid a \wedge Px)}$	$\lambda v.\lambda\psi.[\![\neg]\!]([\![\exists]\!]([\![x]\!], [\![\wedge]\!]([\![R]\!]([\![x]\!], v), \psi)))$	Match (2)
8.	$\boxed{)\ S \wedge \mid \wedge Px)}$	$\lambda\psi.[\![\neg]\!]([\![\exists]\!]([\![x]\!], [\![\wedge]\!]([\![R]\!]([\![x]\!], [\![a]\!]), \psi)))$	Match (2)
9.	$\boxed{)\ S \mid Px)}$	$\lambda\psi.[\![\neg]\!]([\![\exists]\!]([\![x]\!], [\![\wedge]\!]([\![R]\!]([\![x]\!], [\![a]\!]), \psi)))$	Match (1)
10.	$\boxed{)\ T \mid x)}$	$\lambda v.[\![\neg]\!]([\![\exists]\!]([\![x]\!], [\![\wedge]\!]([\![R]\!]([\![x]\!], [\![a]\!]), [\![P]\!](v))))$	Predict & Connect (2)
11.	$\boxed{) \mid)}$	$[\![\neg]\!]([\![\exists]\!]([\![x]\!], [\![\wedge]\!]([\![R]\!]([\![x]\!], [\![a]\!]), [\![P]\!]([\![x]\!]))))$	Match (2)
12.	$\boxed{\mid}$	$[\![\neg]\!]([\![\exists]\!]([\![x]\!], [\![\wedge]\!]([\![R]\!]([\![x]\!], [\![a]\!]), [\![P]\!]([\![x]\!]))))$	Match (1)

APPENDIX B

Languages with Generalized Quantifiers and Restricted Variables

In Chapters 4 and 5, we often tried to elucidate the truth-conditions of possible utterances of English sentences using translations into a first-order language with generalized quantifiers and restricted variables. Some simplified examples are as follows:

(57) a. `Most students`^{1,2} `failed the final exam.`

b. [MOST $\mathtt{v}_1$: STUDENT $\mathtt{v}_1$] FAIL $\mathtt{v}_1$

(58) a. `Most students`^{1,2} `failed the final exam. They`$_2$ `didn't study.`

b. [MOST $\mathtt{v}_1$: STUDENT $\mathtt{v}_1$] FAIL $\mathtt{v}_1 \wedge \neg$STUDY $\mathtt{v}_2$

(59) a. `Alice`^{1} `gave a drawing`^{2,3} `to every teacher`^{4}`. They`$_4$ `each put it`$_3$ `on the wall.`

b. [EVERY $\mathtt{v}_4$: TEACHER $\mathtt{v}_4$][SOME $\mathtt{v}_2$: DRAWING $\mathtt{v}_2$] GAVE $\mathtt{v}_1\mathtt{v}_2\mathtt{v}_4 \wedge$ HANG $\mathtt{v}_4\mathtt{v}_3$

The discussion of the intended interpretation of this language was mostly informal and kept to a bare minimum. This appendix outlines some of the formal details omitted in those chapters and walks through some of these examples more slowly.

In order to give a more precise statement of the semantics of this language, we first need to give an explicit statement of its syntax. We assume, as usual, that we have some fixed stock of predicate symbols $\mathcal{P}$ equipped with an arity function ρ, and a fixed countably infinite stock of variables $\mathcal{V}$. To simplify things, we omit constant symbols in favor of variables interpreted

with singleton restrictions. In addition to the standard propositional connectives $\wedge$ and $\neg$, we will also assume that we have some stock of quantifier symbols $\mathcal{Q}$. Formulas are then straightforwardly generated by the following context-free grammar:

Grammar of First-order Logic with Generalized Quantifiers:

$$T \longmapsto t \ (t \in \mathcal{V})$$
$$Pred_n \longmapsto P \ (P \in \mathcal{P} \text{ and } \rho(P) = n)$$
$$Quant \longmapsto Q \ (Q \in \mathcal{Q})$$
$$QP \longmapsto [Quant\ T\ :\ S]$$
$$S \longmapsto Pred_n \underbrace{T \ldots T}_{n\text{-times}} \mid \neg S \mid (S \wedge S) \mid QP\ S$$

We continue to identify models $\mathcal{M} = \langle U, I \rangle$ as a pair of a universe U and interpretation function I. But in addition to having I map each symbol $P \in \mathcal{P}$ to some $\rho(P)$-ary function over U, we also now assume that it supplies suitable interpretations to the symbols in $\mathcal{Q}$ and $\mathcal{V}$, as well. Quantifier symbols are interpreted by type $\langle 1, 1 \rangle$ generalized quantifiers over U, which are just subsets of $\wp(U) \times \wp(U)$. Some typical examples are as follows:

$$\begin{array}{lll}
\langle A, B \rangle \in EVERY & \text{iff} & A \subseteq B \\
\langle A, B \rangle \in SOME & \text{iff} & A \cap B \neq \emptyset \\
\langle A, B \rangle \in ATLEAST3 & \text{iff} & |A \cap B| > 2 \\
\langle A, B \rangle \in MOST & \text{iff} & |A \cap B| > |A \setminus B| \\
\langle A, B \rangle \in EXACTLYHALF & \text{iff} & |A \cap B| = |A \setminus B|
\end{array}$$

All of our specific examples will use quantifiers drawn from the above list.

Unfortunately, the interpretation of variables will be a bit more complicated to specify. We abstractly model the meaning of a variable as a function from possible states to underlying elements of the universe U. So, in general, we are assuming that it is possible to define a suitable state-space $\mathcal{S}$ over fixed universe U rich enough for our purposes. Note that the standard semantics for first-order languages takes the relevant state-space to be given by U^ω

and identifies the range of possible variable-meanings with projections. We can capture two ideas that are not directly representable in standard formalisms by lifting this assumption. First, variables may be restricted to range only over some proper subset of U. Second, collections of variables may have structural dependencies between each other, in the sense that they cannot jointly assume certain sequences of values even though those values are included among what they each individually range over. To smooth over the presentation of an important definition to follow, it turns out that it will also be somewhat convenient to think of the meanings of variables as potentially being partial, in the sense that they fail to assume a specific value in some states. We model this by including a special value $*$, which a variable assumes whenever it is informally understood as being undefined. This allows us to avoid some of the well-known headaches that stem from partialization. In particular, we can still treat $*$ as though it were otherwise like any other element and adopt the semantic convention that any lexical predication involving it is automatically true (rather than neither true nor false).

Given a variable-meaning v, let $def(v) = \{\sigma \in \mathcal{S} : v(\sigma) \neq *\}$. We say that $v_{i_1}...v_{i_n}$ represents R just in case $def(v_{i_j}) = def(v_{i_k})$ for all $j, k \leq n$ and $\{\langle a_1, \ldots, a_n\rangle \mid \exists\sigma.\ v_{i_1}(\sigma) = a_1 \,\&\, \ldots \,\&\, v_{i_n}(\sigma) = a_n \neq *\} = R$. Our main desideratum is that $\mathcal{S}$ is rich enough to represent every n-ary relation over U. Thankfully, we can straightforwardly construct a state-space with required properties by a relatively simple generalization of the standard approach. In particular, we will still identify variable-states over U with sequences of elements of U and take variable-meanings to be the possible projections of those sequences. First, define a sequence of elements from a universe U of length α (where α is any ordinal) to be a map from α to U. Given two sequences $\vec{a}$ and $\vec{b}$ of length α and β respectively, their concatenation $\vec{a}^\frown\vec{b}$ is the sequence of length $\alpha + \beta$ such that for any $\gamma < \alpha$, $\vec{a}^\frown\vec{b}(\gamma) = \vec{a}(\gamma)$, and for any $\alpha \leq \alpha + \gamma' < \alpha + \beta$, $\vec{a}^\frown\vec{b}(\alpha + \gamma') = \vec{b}(\gamma')$. This operation can obviously be lifted to sets of sequences R and S, so that their concatenation $R^\frown S$ is $\{\vec{a} \frown \vec{b} \mid \vec{a} \in R, \vec{b} \in S\}$. Then just note that since the set of finite relations over $U \cup \{*\}$ is well-ordered by some ordinal

κ, we can arrive at a set of sequences with the desired properties by iterating this lifted concatenation operation along κ and taking unions at limit stages. So we may now officially take $\mathcal{S}$ to be the end result of that process and require that the interpretation function I of $\mathcal{M}$ to map each variable in $\mathcal{V}$ to some ordinal indexed projection π_α, where $\alpha < lh(\sigma)$ for any $\sigma \in \mathcal{S}$. Note that since we assume U^n is representable for every n, we may consider $\mathcal{M}^{Stand}$ (the *standardization* of $\mathcal{M}$) for any model $\mathcal{M}$, which is a model that is otherwise like $\mathcal{M}$ except that the interpretation of every variable has been chosen so that it jointly represents U^ω (as it does in the standard case). And more locally, we can also standardize individual variables t by considering $\mathcal{M}_t^{Stand}$, which is a model otherwise like $\mathcal{M}$ except that $I(t)$ represents U and for all v that represent P, $I(t)$ and v jointly represent $U \times P$.

With our notion of model now in place, we can give our official definition of truth with the help of a few more auxiliary definitions. The denotation of a term t with respect to a state σ and model $\mathcal{M}$ ($den_\sigma^{\mathcal{M}}(t)$) is just $I(t)(\sigma)$. Given a variable-meaning v and states σ and σ', we say $\sigma \sim_v \sigma'$ (σ is a v-variant of σ') iff $v = \pi_\alpha$ and $\sigma(\beta) = \sigma'(\beta)$ for all ordinals $\beta \neq \alpha$. We define $range(v) = \{a \mid \exists \sigma . v(\sigma) = a \ \& \ a \neq *\}$, and as a slight abuse of notation, we also write $range^{\mathcal{M}}(t) = range(I(t))$. Finally, the set of (defined) possible values of the term t that satisfy ϕ with parameters σ is given by $\| \phi \|_t^{\mathcal{M},\sigma} = I(t)[\{\sigma' \mid \sigma' \sim_{I(t)} \sigma \ \& \ \mathcal{M}, \sigma' \vDash \phi \ \& \ I(t)(\sigma') \neq *\}]$. The recursive definition of truth with respect to a variable-state then runs as follows:

Definition of Truth:

$$
\begin{array}{lll}
\mathcal{M}, \sigma \vDash Pt_1 \ldots t_n & \text{iff} & \langle den_\sigma^{\mathcal{M}}(t_1), \ldots, den_\sigma^{\mathcal{M}}(t_n) \rangle \in I(P) \text{ or} \\
 & & den_\sigma^{\mathcal{M}}(t_i) = * \text{ for some } i \leq n \\
\mathcal{M}, \sigma \vDash \neg\phi & \text{iff} & \mathcal{M}, \sigma \nvDash \phi \\
\mathcal{M}, \sigma \vDash (\phi \wedge \psi) & \text{iff} & \mathcal{M}, \sigma \vDash \phi \text{ and } \mathcal{M}, \sigma \vDash \psi \\
\mathcal{M}, \sigma \vDash [Q \ t : \phi] \ \psi & \text{iff} & range^{\mathcal{M}}(t) = \| \phi \|_t^{\mathcal{M}_t^{Stand},\sigma} \text{ and} \\
 & & \langle range^{\mathcal{M}}(t), \| \psi \|_t^{\mathcal{M},\sigma} \rangle \in I(Q)
\end{array}
$$

Finally, we say a formula ϕ is true (simpliciter) with respect to $\mathcal{M}$ ($\mathcal{M} \vDash \phi$) just in case it

is true with respect to every $\sigma \in \mathcal{S}$.

Notice that the way we have implemented the clause for the truth-conditions of formulas involving generalized quantifiers is quite different from how it is normally done. This is because our account placed a much greater amount of emphasis on the role that speakers' communicative intentions play in determining the contents of their utterances. Our basic idea was that when this formal language is used to represent the content of a natural language utterance, *all* of the variables occurring in that representation (including bound ones) need to be thought of as being somehow supplied by those intentions. In particular, when a speaker uttered a quantifier phrase, it was assumed that they had some variable-meaning in mind that represents the suitable domain of quantification. To partly capture this idea, we have implemented the restrictive part of the generalized quantifier clause as a kind of check. Normally, this clause is defined in a way that constrains our attention to only the values of an unrestricted variable that satisfy the explicitly supplied restricting element. Our clause, on the other hand, checks that the value range of a contextually restricted variable coincides with the set that is denoted by a restrictive element. In other words, we treat it as a truth-conditional effect that the domain variable intended by the speaker has the properties that are explicitly ascribed to it by their utterance.

With these preliminaries in place, we can now walk through the examples that this appendix began with in a bit more detail. First, consider the simplest example (57). We noted that when a speaker utters (57a), she is normally expected to have some group of students in mind that contextually supplies the required domain of quantification. The formalization in (57b) directly captures this intuition by using a variable $\mathtt{v}_1$ that we interpret as ranging over all and only the contextually relevant students. If we also interpret the predicate STUDENT as applying only to those students, the check on the range of the variable imposed by the quantifier MOST is guaranteed to succeed. The correct truth-conditions for the utterance are therefore obtained if we also interpret the remaining pieces of the formalization in the obvious way – i.e., MOST by the generalized quantifier $MOST$ and the

predicate FAIL as applying to all and only the students who failed.

Our explanation of (58) begins in exactly the same way, but we now need to additionally specify how the second variable $\mathtt{v}_2$ is interpreted. In addition to the domain of quantification captured by $\mathtt{v}_1$, we noted that a speaker of (58a) would also normally be expected to have some subregion of that domain in mind that suffices to establish the truth of her claim. The formalization in (58b) tries to capture that idea by having the interpretation of $\mathtt{v}_2$ be determined by those communicative intentions and so suitably related to the interpretation of $\mathtt{v}_1$. One absolutely minimal constraint is the requirement that the one to include the other, in the sense that $range^{\mathcal{M}}(\mathtt{v}_2) \subseteq range^{\mathcal{M}}(\mathtt{v}_1)$. But as it turns out, this will be a little too weak to handle certain other more complicated examples that we would like our theory to treat. In essence, the issue is that there may be a situation such that $range^{\mathcal{M}}(\mathtt{v}_1) \subseteq range^{\mathcal{M}}(\mathtt{v}_2)$ but also some further variable $\mathtt{v}_3$ and pair of values $\langle a_1, a_2\rangle$ such that $\langle \mathtt{v}_2, \mathtt{v}_3\rangle$ but not $\langle \mathtt{v}_1, \mathtt{v}_3\rangle$ may assume $\langle a_1, a_2\rangle$ in a single state despite the fact that $a_1 \in range^{\mathcal{M}}(\mathtt{v}_1)$. To get around this problem, we would like the relevant notion of inclusion to respect how denotations are allowed to vary from state to state. The structured inclusion of variable-meanings is therefore defined as follows: $v \sqsubseteq v'$ iff $def(v) \subseteq def(v')$ and for all $\sigma \in def(v)$: $v(\sigma) = v'(\sigma)$. But even the requirement that $I(\mathtt{v}_1) \sqsubseteq I(\mathtt{v}_2)$ is not quite enough in this case. In particular, we must also capture the idea that the interpretation of $\mathtt{v}_2$ is a *witness* to the quantifier that ranges over the domain given by $\mathtt{v}_1$. In general, what counts as a contextually acceptable witness cannot be motivated on entirely semantic grounds since that is a partly conventional matter, and so we will instead have to alternatively stipulate an extra condition for each English quantifier word being symbolized. For our purposes, the following three suffice:

$$
\begin{array}{lll}
wit_{\mathtt{every}}(v_1, v_2) & \text{iff} & v_1 = v_2 \\
wit_{\mathtt{a}}(v_1, v_2) & \text{iff} & v_1 \sqsubseteq v_2 \text{ and } |range(v_1)| = 1 \\
wit_{\mathtt{most}}(v_1, v_2) & \text{iff} & v_1 \sqsubseteq v_2 \text{ and } |range(v_1)| > |range(v_2) \setminus range(v_1)|
\end{array}
$$

We can round out our explanation of (58b), then, by requiring $wit_{\mathtt{most}}(I(\mathtt{v}_2), I(\mathtt{v}_1))$ and that $\mathcal{M} \vDash$ FAIL $\mathtt{v}_2$. The desired truth-conditions are then yielded, because our definition of truth

(simpliciter) universally quantifies over states, and so the free variable $\mathtt{v}_2$ is treated as being implicitly universally quantified. In other words, (58b) is true just in case the failing students that the speaker has in mind constitute more than half of the totally of contextually relevant students, and that each of them did not study. Note that the conditions on $\mathtt{v}_2$ are stated as meta-language constraints on its acceptable interpretations rather than being directly represented as an object-language condition in the representation in (58b). This reveals that we prefer to treat them as being mutually presupposed by the conversational participants rather than as something that the speaker directly asserts. In other words, it is a (defeasible) precondition on the speaker saying anything at all by her second utterance that she have in mind most of the contextually relevant students and that those students she has in mind all failed. Of course, speaker reference may still succeed when the presuppositions are not met. But some corrective measure is normally called for in those cases.

Explaining the dependent anaphora exemplified by (59) is unsurprisingly a bit more complicated. Here, the witness that our account says a speaker must have in mind when uttering (59a) is to the polyadic quantifier that arises by scoping `every` over `a`. In cases like this, we suggested that the conventional constraints on the acceptable witness of the overall quantifier arise from the conventional constraints on its individual components. We formalize this idea in our present framework with the following definitions. Given variable-meanings v_1 and v_2 and a value $a \in \mathit{range}(v_1)$, v_3 is a distributive slice of v_2 over v_1 with respect to a ($v_3 \in [v_2/^a v_1]$) iff $v_3 \sqsubseteq v_2$ and $\mathit{def}(v_3) = \{\sigma : v_1(\sigma) = a\}$. Using this notion, the conventional constraints on the witnesses for an iterated quantifier arising from scoping the expression Q_1 over Q_2 can be schematically specified in terms of the constraints that apply to Q_1 and Q_2 individually:

$$\mathit{itWit}_{Q_1 \cdot Q_2}(v_1, v_2, v_3, v_4) \quad \text{iff} \quad \mathit{wit}_{Q_1}(v_1, v_2) \text{ and for all } a \in \mathit{Range}(v_1),\ \mathit{wit}_{Q_2}(v_5, v_6)$$
$$\text{for every } v_5 \in [v_3/^a v_1] \text{ and } v_6 \in [v_4/^a v_1]$$

With this in place, we can now walk through our representation of the truth conditions of a typical utterance of (59a) given in (59b). This representation involves four variables. The

variable $\mathtt{v}_1$ is associated with the occurrence of the proper name `Alice`, and so is naturally interpreted as ranging only over the individual designated by that name. The variables $\mathtt{v}_2$ and $\mathtt{v}_3$ are associated with the occurrence of the indefinite `a drawing`, and respectively track the domain and witness elements the speaker must have had in mind. Finally, since the quantifier designated by `every teacher` has a unique witness that coincides with its domain, we can represent both by the same variable $\mathtt{v}_4$. If the other expressions are all given their obvious interpretations, then the truth-conditions for the first conjunct of (59b) are what they need to be – it is true just in case every teacher is such that there is some drawing that Alice gave them. We then treat as presupposed in the conversation that $itWit_{\mathtt{every\cdot a}}(I(\mathtt{v}_4), I(\mathtt{v}_4), I(\mathtt{v}_3)I(\mathtt{v}_2))$ and that $\mathcal{M} \vDash \textsc{gave}\ \mathtt{v}_1\mathtt{v}_3\mathtt{v}_4$. If we unpack the definition just given a little bit, we see that this is the (trivial) requirement that $I(\mathtt{v}_4) = I(\mathtt{v}_4)$, and that for any $a \in range(I\mathtt{v}_4)$, $v \in [\mathtt{v}_3/^a\mathtt{v}_4]$, and $v' \in [\mathtt{v}_2/^a\mathtt{v}_4]$, $v \sqsubseteq v'$ and $|range(v)| = 1$. With enough squinting, this requirement can, in this context, be seen to be equivalent to the claim that the pair of variables $\langle \mathtt{v}_4, \mathtt{v}_3 \rangle$ jointly represents a function mapping each contextually relevant teacher to some contextually relevant drawing given to them by Alice. The truth-conditions for the second conjunct then come out as they need to be, since the open formula $\mathcal{M} \vDash \textsc{hang}\ \mathtt{v}_4\mathtt{v}_3$ iff $\mathcal{M}, \sigma \vDash \textsc{hang}\ \mathtt{v}_4\mathtt{v}_3$ for every $\sigma \in \mathcal{S}$, and $\langle \mathtt{v}_3, \mathtt{v}_4 \rangle$ are presupposed to only jointly assume pairs of values $\langle a, b \rangle$ such that b was given a.

APPENDIX C

Strongly Compositional Dynamics

In Chapter 5, we suggested that it was possible to give all pronoun occurrences a uniform treatment:

(60) `Alice`^{1} `brought her`$_1$ `bluebook.`

(61) `Alice brought [points]`^{1} `her`$_1$ `bluebook.`

(62) `Every student`^{1} `brought her`$_1$ `bluebook.`

The idea was that *deictic* and *bound* occurrences (exemplified by (61) and (62) could be assimilated to *anaphoric* occurrences (exemplified by (60)). Roughly put, the content of an anaphoric pronoun in context is always the value of the contextually most salient discourse referent. The one hiccup with this analysis is that certain terms *introduce* new discourse referents as soon as they are encountered. And systematically accommodating this aspect of their semantic behavior looked like it may run afoul of strong compositionality, which is the most distinctive aspect of the Kaplanian interpretive architecture. But we maintained that this problem is merely apparent and that so long as the contents of expressions were computable *on demand*, there is no real threat. This appendix outlines a rudimentary method of formally implementing this suggestion as a basic proof of concept.

First, we will need to say a little bit about the shape of our lexicon. We will obviously need to continue to follow Kaplan in distinguishing between a context-invariant rule-like notion of meaning (character) and the contextually saturated results of applying those rules on a given occasion (content). Therefore, one component of our proposal will associate a function

from contexts of utterance to contents to each lexical item, as is normal. Additionally, any expression capable of serving as the antecedent to a pronoun will also need to be assigned a function from contexts to contexts, which will allow us to model the addition of new contextual information when those expressions are encountered. To simplify our presentation a bit, we can avoid the need to supply two separate lexical entries for certain expressions by instead associating each lexical item with a single function from a context c to *pairs* $\langle x, c' \rangle$, where x is the expression's content in c and c' is the result of adding new information about potential antecedents to c.

We will unsurprisingly also need to take a modular approach to meaning composition. Since our account needs to adhere to strong compositionality, we will need a family of *content composition operators* for building larger contents out of smaller ones. Following standard assumptions, we may take these to be adequately characterized by the single rule of *type-driven function application*. In informal terms, this rule says that local syntactic combination should be semantically interpreted as a function being saturated by its argument whenever possible. And in order to model the growth of contextual information as a text is read in left-to-right order, we will additionally need to define a family of *linear evaluation operators*. Informally speaking, these operators compute the content of a complex expression $e_1^\frown e_2$ in c by first computing the contents of e_1 and e_2 (in that order) and then combining them using some contextually acceptable content composition operator. These will allow us to formally recognize that acts of speaker reference produce context-changing side effects that are significant for the interpretation of anaphora. In particular, the fact that e_2 is evaluated after e_1 will mean that entities referred to in the production of e_1 will be available to serve as potential contents for any pronouns occurring in e_2.

Following Shan (2007) and Unger (2012), we can take some ideas from the semantics of programming languages to help formally implement these ideas. This work starts with the framework presented in Moggi (1991), which offers a very simple basis for developing the semantics of programming languages with side-effects. Conceptually speaking, it makes an

intuitive distinction between a *value* of some type and a *computation* that yields a value of some type as well as possibly producing additional computational side-effects when executed. What additional effects are produced can vary depending on the notion of computation being modeled. The central insight is that a wide variety of different notions of computation can naturally be thought of as sharing a common core behavior, which a very small number of axioms can describe. Moreover, those axioms turn out to be the same as those governing the category-theoretic concept of a *monad*, thus providing us an extremely useful mathematical structure that can be used to model certain features of programming languages. For our current purposes, we can safely work with the following simplified type-theoretic definition:

> (**Monads**) A *monad* is a triple $\langle \Diamond, \eta, \star \rangle$ of a type constructor $\Diamond$, a unary function $\eta : \forall\alpha.\alpha \to \Diamond\alpha$, and a binary function $\star : \forall\alpha\forall\beta.\Diamond\alpha \to (\alpha \to \Diamond\beta) \to \Diamond\beta$, subject to the laws:

$$\eta(x) \star \lambda y.f \equiv f[x/y]$$
$$m \star \lambda x.\eta(x) \equiv m$$
$$m \star \lambda x.(f \star \lambda y.g) \equiv (m \star \lambda x.f) \star \lambda y.g \ (x \notin FV(g))$$

The intuition behind this definition is that we begin by imagining ourselves in a purely extensional context. Two programs are identified just in case they share the same input/output behavior on an antecedently given space of values. In order to make finer-grained distinctions among programs, we then enrich the space to include computations that encode some kind of useful operational information over and above information about resulting values. In this new context, two previously indistinguishable programs $f, g : A \to B$ will be associated with new modified programs $f^*, g^* : A \to \Diamond B$, which may now turn out to be denotationally distinct, depending on the extra structure that we are tracking with the computation-type $\Diamond B$. The return operator η gives us a systematic way of relating the initial type A of values and the new type $\Diamond A$ of computations, and the bind operator $\star$ tells us how to sequence

programs now that we acknowledge the possibility of computational side-effects. With this in mind, the axioms just express some natural constraints on how we would like η and $\star$ to behave.

This approach is appropriate for our current purposes mainly because of the high degree of success it has enjoyed in many practical programming contexts. In particular, it has proven extremely useful as a means of systematically embedding features of impure programming languages into otherwise pure ones.[1] This is partly because it yields a straightforward procedure that can systematically identify imperative programs with terms of the typed lambda calculus. Here is a particularly simple example of that procedure in action:

```
f := expr1
x := expr2      ⟹  expr1 ⋆ (λf.expr2 ⋆ λx.η(f(x)))
return (f x)
```

This framework therefore allows us to stick to mathematical tools that are already very well-understood and have well-established roles in ordinary semantic inquiry. And for our present purposes, the *state* monad provides us with exactly what we need:

(**State Monad**) The *State* monad is $\langle \Diamond_k, \eta_k, \star_k \rangle$, where:

($\Diamond_k$) For any type τ, $\Diamond_k \tau = k \to (\tau \times k)$
(η_k) For any term x^α, $\eta_k x = \lambda c^k . \langle x, c \rangle$
($\star_k$) For any terms $x^{\Diamond\alpha}$ and $f^{\alpha \to \Diamond\beta}$, $x \star_k f = \lambda c^k . f \pi_1(xc) \pi_2(xc)$

Before we can specify our lexicon and evaluation operators using this monad, we need to say a few words about the basic type-theoretic assumptions that we will be making. To keep unnecessary complexities to an absolute minimum, we have confined ourselves to an entirely extensional language fragment. So we may follow standard practices in assuming

[1] See Wadler (1995) for an extremely accessible introduction to this line of work.

that we have two basic types e and t and that any content has a type generated from just those two using the constructor $\rightarrow$. But while these are ordinarily interpreted as (e)ntities and (t)ruth-values, we will instead assume they are interpreted as variable-meanings and open propositions. In other words, an object of type e will be understood as a function from variable-states to entities, and an object of type t will be understood as a function from variable-states to truth-values. Additionally, we take k to be the type of conversational contexts, which for simplicity, we interpret as the set of finite strings over the alphabet e (i.e., lists of variable-meanings). Given these assumptions, the interpretation of a lexical item whose content in context is something of type τ can, by our earlier remarks, be naturally identified with something of type $\Diamond_k\tau$ – that is, as a function from contexts to pairs of contents of type τ and (updated) contexts. The following lexicon therefore suffices to handle the examples (60)-(62):

Lexicon for Strongly Compositional Dynamics

Expression	Interpretation	Type
`Alice`	$\lambda c^k.\langle v_{\mathbf{a}}, v_{\mathbf{a}}{}^\frown c\rangle$	$\Diamond e$
`student`	$\lambda c^k.\langle v_{\mathbf{std}}, v_{\mathbf{std}}{}^\frown c\rangle$	$\Diamond e$
`bluebook`	$\lambda c^k.\langle v_{\mathbf{blb}}, v_{\mathbf{blb}}{}^\frown c\rangle$	$\Diamond e$
`her`	$\lambda c^k.\langle \lambda x^e.poss(sel(c), x), c\rangle$	$\Diamond(e \rightarrow e)$
`brought`	$\lambda c^k.\langle \lambda x^e \lambda y^e.Brought(y, x), c\rangle$	$\Diamond(e \rightarrow (e \rightarrow t))$
`every`	$\lambda c^k.\langle \lambda x^e \lambda P^{e\rightarrow t}.\forall x.P(x)), c\rangle$	$\Diamond(e \rightarrow (e \rightarrow t) \rightarrow t)$

These entries involve several primitives that I now need to say a bit more about. First, note that proper and common nouns are given uniform interpretations as designating *restricted* variable-meanings. So the proper name `Alice` designates the variable-meaning $v_{\mathbf{a}}$, which is subject to a singular restriction to range over only Alice. And the common nouns `student` and `bluebook` designate the the variable-meanings $v_{\mathbf{std}}$ and $v_{\mathbf{blb}}$, which are respectively restricted to range over the students and the bluebooks.[2] In a way, then, our view constitutes

[2]We are glossing over a technical complexity here. In general, different occurrences of the same common

a partial revival of the Aristotelian notion of a subject (though one that ironically needs to take seriously some of the linguistic innovations that set modern logic apart from its ancient counterparts).[3] These expressions also have non-trivial effects on context. The entities they designate are added to the front of the list of possible pronominal antecedents whenever they are encountered. The possessive pronoun `her`, on the other hand, designates an explicit variable-restricting function. It takes a variable-meaning v and returns the variable-meaning $v' \sqsubseteq v$ restricted to range over all and only the elements of $range(v)$ that are also possessed by the contextually most salient entity in context. The selection of that entity is accomplished by an anaphora resolution mechanism sel, which we model as the function that merely takes the head of a list.[4] The denotations of `brought` and `every` are much simpler to specify since these expressions do not involve any interaction (either reading or writing) with context. They can, therefore, more or less straightforwardly be understood as lifts of their ordinary Tarskian meanings using the return operator η.

We now need to specify a family of evaluation operators to round out our discussion. In particular, we need three to treat the examples we started with, which all share the same basic shape. Two are used to account for local content composition as function application – one for when the argument-designating expression is on the left and one for when the argument-designating expression is on the right. The third is used to accommodate for potential demonstrative updates. These are situations in which expressions are used *only* for their effect on context. In these cases, their contents are discarded rather than directly making

noun will have to designate different variable-meanings with the same domain restrictions so that those expression occurrences do not incur any illicit structural dependencies with each other. This is easily modeled with a slightly more complicated lexical entry – roughly speaking, the content of a common noun in context can be stipulated to be the first *unused* variable-meaning that meets the suitable requirements on its range and dependencies. But since this is not needed to treat the examples we are now examining, we set this issue aside here.

[3]Compare this account with Lasersohn (2021), which presents a very similar view.

[4]This formalizes the idea that the 'most salient' entity in context is the one that was most recently mentioned. This is obviously a grotesque oversimplification of this problem, which is among the most difficult problems in natural language processing.

semantic contributions. In particular, this is how we will choose to analyze the significance of the gesture in (61). More specifically, the interpretive rule governing the significance of the speaker's pointing can be formally modeled as an object of type $\Diamond_k e$, like other potentially subject-designating expressions are. But rather than making a *direct* semantic contribution in context (i.e., providing a content that is integrated into the running meaning estimate), it makes an *indirect* contribution by promoting its content in contextual salience. This then allows the subsequent pronoun to access it and so serves to help determine the content of that latter expression occurrence instead:

Linear Evaluation Operators for Strongly Compositional Dynamics

Operator	Specification		
$\oslash$	$m_1^{\Diamond_k(\alpha\to\beta)} \oslash m_2^{\Diamond_k\alpha}$	$=$	$m_1 \star_k (\lambda f.m_2 \star_k \lambda x.\eta_k(f(x)))$
$\obslash$	$m_2^{\Diamond_k\alpha} \obslash m_1^{\Diamond_k(\alpha\to\beta)}$	$=$	$m_2 \star_k (\lambda x.m_1 \star_k \lambda f.\eta_k(f(x)))$
$\oplus$	$m_1^{\Diamond_k\alpha} \oplus m_2^{\Diamond_k\beta}$	$=$	$m_1 \star_k (\lambda x.m_2 \star_k \lambda y.\eta_k(y))$

We use $\oslash$ to interpret any local syntactic configuration where the semantic argument is on the left, $\obslash$ to interpret any local syntactic configuration where the semantic argument is on the right, and $\oplus$ to perform a demonstrative update. Finally, we assume the semantic rule governing demonstrative gestures can be modeled as $point_{\mathbf{that}} = \lambda c^k.\langle v_{\mathbf{that}}, v_{\mathbf{that}}{}^\frown c\rangle$.[5] Putting everything together, we get the following results:

(60′) a. `Alice`[1] `brought her`$_1$ `bluebook.`

b. $[\![\texttt{Alice}]\!] \obslash ([\![\texttt{brought}]\!] \oslash ([\![\texttt{her}]\!] \oslash [\![\texttt{bluebook}]\!]))$

c. $Brought(v_{\mathbf{a}}, poss(v_{\mathbf{a}}, v_{\mathbf{blb}}))$

d. $\langle v_{\mathbf{blb}}, v_{\mathbf{a}}\rangle$

(61′) a. `Alice brought [points]`[1] `her`$_1$ `bluebook.`

[5]This is obviously an oversimplification – the significance of gesture is clearly bidirectionally context-sensitive in that it both reads its value from context *and* promotes it value in salience. This is very easy to model but requires using a richer notion of context than the one we have been working with, and so we have only modeled them as being context-sensitive in the one direction that matters for our example.

b. $[\![\texttt{Alice}]\!] \oslash ([\![\texttt{brought}]\!] \oslash (point_{\mathbf{that}} \oplus ([\![\texttt{her}]\!] \oslash [\![\texttt{bluebook}]\!])))$

c. $Brought(v_{\mathbf{a}}, poss(v_{\mathbf{that}}, v_{\mathbf{blb}}))$

d. $\langle v_{\mathbf{blb}}, v_{\mathbf{that}}, v_{\mathbf{a}} \rangle$

(62′) a. `Every student`^{1} `brought her`$_1$ `bluebook.`

b. $([\![\texttt{every}]\!] \oslash [\![\texttt{student}]\!]) \oslash ([\![\texttt{brought}]\!] \oslash ([\![\texttt{her}]\!] \oslash [\![\texttt{bluebook}]\!]))$

c. $\forall v_{\mathbf{std}}.Brought(v_{\mathbf{std}}, poss(v_{\mathbf{std}}, v_{\mathbf{blb}}))$

d. $\langle v_{\mathbf{blb}}, v_{\mathbf{std}} \rangle$

In each of the above, the first entry gives the natural language expression to be analyzed, the second entry gives the unreduced analysis that our framework yields, the third entry gives the resulting content when these expressions are evaluated against the null-context $\langle\rangle$, and the fourth entry gives the resulting output context under those same conditions.

Bibliography

Abbott, B. 2003. A reply to szabó's "descriptions and uniqueness". *Philosophical Studies*, 113(3):223–231.

Altmann, G. and Steedman, M. 1988. Interaction with context during human sentence processing. *Cognition*, 30:191–238.

Brasoveanu, A. 2008. Donkey pluralities: plural information states versus non-atomic individuals. *Linguistics and philosophy*, 31(2):129–209.

Brasoveanu, A. 2010. Structured anaphora to quantifier domains. *Information and Computation*, 208(5):450–473.

Breckenridge, W. and Magidor, O. 2012. Arbitrary reference. *Philosophical Studies*, 158(3):377–400.

Bresnan, J. and Kaplan, R. 1982. Grammars as mental represenations of language. In Bresnan, J., editor, *The Mental Representation of Grammatical Relations*. MIT Press.

Büring, D. 2005. *Binding theory*. Cambridge University Press.

Cresswell, M. J. 1990. *Entities and indices*, volume 41. Springer.

Cumming, S. 2015. The dilemma of indefinites. In Bianchi, A., editor, *On Reference*. Oxford University Press.

Davidson, D. 1965. Theories of meaning and learnable languages. In Bar-Hillel, Y., editor, *Proceedings of the International Congress for Logic, Methodology, and Philosophy of Science*. North-Holland.

Davies, M. and Humberstone, L. 1980. Two notions of necessity. *Philosophical Studies*, 38(1):1–30.

Dekker, P. 2000. Coreference and representationalism. In *Reference and anaphoric relations*, pages 287–310. Springer.

Dekker, P. 2002. Meaning and use of indefinite expressions. *Journal of Logic, Language and Information*, 11(2):141–194.

Dever, J., Pickel, B., and Rabern, B. 2018. Reviving the parameter revolution in semantics. In Ball, D. and Rabern, B., editors, *The Science of Meaning*. Oxford University Press.

Donnellan, K. S. 1966. Reference and definite descriptions. *The philosophical review*, 75(3):281–304.

Dummett, M. 1991. *The Logical Basis of Metaphysics*. Harvard University Press.

Elbourne, P. 2013. *Definite descriptions*, volume 1. Oxford University Press.

Elworthy, D. A. 1995. A theory of anaphoric information. *Linguistics and Philosophy*, pages 297–332.

Evans, G. 1977. Pronouns, quantifiers, and relative clauses (i). *Canadian journal of philosophy*, 7(3):467–536.

Fiengo, R. and May, R. 1994. *Indices and identity*. MIT press.

Fiengo, R. and May, R. 2006. De lingua belief. cambridge ma: Bradford book.

Fine, K. 2003. The role of variables. *Journal of Philosophy*, 100(12):605–631.

Fine, K. 2007. *Semantic Relationism*. Blackwell Publishing.

Fodor, J. A. 1975. *The language of thought*, volume 5. Harvard university press.

Fodor, J. A. and Lepore, E. 1992. Holism: A Shopper's Guide. Blackwell.

Fodor, J. D. and Sag, I. A. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy*, 5:355–398.

Frege, G. 1884. *Die Grundlagen der Arithmetik*. Breslau.

Frege, G. 1892. über sinn und bedeutung. *Zeitschrift für Philosophie Und Philosophische Kritik*, 100(1):25–50.

Frege, G. 1893. *Grundgesetze der Arithmetik*, volume 1. Pohle. English translation in P. Ebert and M. Rossberg (editors and translators), *Gottlob Frege: Basic Laws of Arithmetic*. Oxford University Press (2013).

Geach, P. 1980. *Reference and Generality: An Examination of Some Medieval and Modern Theories*. Cornell University Press, third edition.

Geach, P. T. 1967. Intentional identity. *The Journal of Philosophy*, 64(20):627–632.

Glanzberg, M. 2009. Semantics and truth relative to a world. *Synthese*, 166(2):281–307.

Glanzberg, M. 2011. More on operators and tense. *Analysis*, 71(1).

Glanzberg, M. and King, J. 2020. Binding, compositionality, and semantic values. *Philosophers' Imprint*, 19(2).

Grice, P. 1975. Logic and conversation. In Davidson, D. and Harman, G., editors, *The Logic of Grammar*. Dickenson.

Groenendijk, J. and Stokhof, M. 1991. Dynamic predicate logic. *Linguistics and philosophy*, pages 39–100.

Grune, D. and Jacobs, C. J. 2008. *Parsing Techniques*. Springer-Verlag, second edition.

Hacking, I. 1979. What is logic? *The Journal of Philosophy*, 76(6):285–319.

Haddock, N. J. 1989. Computational models of incremental semantic interpretation. *Language and Cognitive Processes*, 4(3):337–368.

Harel, D., Meyer, A. R., and Pratt, V. R. 1977. Computability and completeness in logics of programs (preliminary report). In *Proceedings of the ninth annual ACM symposium on Theory of computing*, pages 261–268.

Hawthorne, J. and Manley, D. 2012. *The reference book*. Oxford University Press.

Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts, Amherst.

Henkin, L., Monk, J. D., and Tarski, A. 1971. *Cylindric Algebras*, volume 64 of *Studies in Logic and The Foundations of Mathematics*. North-Holland Publishing Company.

Hindley, J. R. and Seldin, J. P. 2008. *Lambda-Calculus and Combinators: an Introduction*. Cambridge Univ Press.

Humberstone, L. 2000. What *Fa* says about *a*. *Dialectica*, 54(1):3–28.

Iacona, A. 2018. *Logical Form: Between Logic and Natural Language*. Springer Verlag.

Jacobson, P. 1999. Towards a variable-free semantics. *Linguistics and Philosophy*, 22(2):117–184.

Jacobson, P. 2003. Binding without pronouns (and pronouns without binding). In Kruijff, G.-J. M. and Oehrle, R. T., editors, *Resource-Sensitivity, Binding and Anaphora*. Springer, Dordrecht.

Janssen, T. 1997. *Handbook of Logic and Language*, chapter Compositionality. North-Holland.

Kamp, H. 1971. Formal properties of 'now'. *Theoria*, 37(3):227–273.

Kamp, H. 1981. A theory of truth and semantic representation. In Groenendijk, J., Janssen, T., and Stokhof, M., editors, *Formal Methods in the Study of Language*. Dordrecht: Foris.

Kamp, H. 1990. Prolegomena to a structural account of belief and other attitudes. *Current Research in the Semantics/Pragmatics Interface*, page 513.

Kaplan, D. 1986. Opacity. In Hahn, L. E. and Schilpp, P. A., editors, *The Philosophy of W. V. Quine*. Open Court.

Kaplan, D. 1989a. Afterthoughts. In Almog, J., editor, *Themes from Kaplan*, pages 565 – 614. Oxford University Press.

Kaplan, D. 1989b. Demonstratives. In Almog, J., editor, *Themes from Kaplan*, pages 481 – 563. Oxford University Press.

Karttunen, L. 1976. Discourse referents. In *Notes from the linguistic underground*, pages 363–385. Brill.

King, J. C. 2003. Tense, modality, and semantic values. *Philosophical perspectives*, 17:195–245.

Klein, U. and Sternefeld, W. 2017. Same same but different: An alphabetically innocent compositional predicate logic. *Journal of Philosophical Logic*, 46:65–95.

Kracht, M. 2011. *Interpreted languages and compositionality*, volume 89. Springer.

Kreisel, G. and Krivine, J. 1967. *Elements of Mathematical Logic*. Studies in Logic and the Foundations of Mathematics. North-Holland.

Krifka, M. 1996. Parametrized sum individuals for plural anaphora. *Linguistics and philosophy*, pages 555–598.

Kripke, S. 1977. Speaker's reference and semantic reference. *Midwest studies in philosophy*, 2(1):255–276.

Landman, F. 1986. Pegs and alecs. In *Proceedings of the 1986 Conference on Theoretical aspects of reasoning about knowledge*, pages 45–61.

Lasersohn, P. 2021. Common nouns as modally non-rigid restricted variables. *Linguistics and Philosophy*, 44(2):363–424.

Lasnik, H. 1976. Remarks on coreference. *Linguistic analysis*, 2.

Lavine, S. 2000. Quantification and ontology. *Synthese*, 124(1):1–43.

Lepore, E. 1983. What model theoretic semantics cannot do? *Synthese*, pages 167–187.

Lewis, D. 1980. Index, context, and content. In *Philosophy and grammar*, pages 79–100. Springer.

Lewis, D. K. 1975. Adverbs of quantification. In Keenan, E. L., editor, *Formal Semantics of Natural Language*, pages 178–188. Cambridge University Press.

Marcus, R. B. 1972. Quantification and ontology. *Noûs*, pages 240–250.

Moggi, E. 1991. Notions of computation and monads. *Information and control*, 93(1):55–92.

Monk, J. D. 1976. *Mathematical Logic*. Graduate Texts in Mathematics. Springer-Verlag.

Montague, R. 1973. The proper treatment of quantification in ordinary english. In Suppes, P., Moravcsik, J., and Hintikka, J., editors, *Approaches to Natural Language*. Dordrecht.

Muskens, R. 1996. Combining montague semantics and discourse representation. *Linguistics and Philosophy*, 19(2):143–186.

Neale, S. 1990. *Descriptions*. MIT Press.

Nouwen, R. 2007. On dependent pronouns and dynamic semantics. *Journal of Philosophical Logic*, 36(2):123–154.

Pagin, P. and Westerståhl, D. 2010. Compositionality ii: Arguments and problems. *Philosophy Compass*, pages 265–282.

Partee, B. H. and Marsh, W. 1987. How non-context free is variable binding? In Savitch, W. J., Bach, E., Marsh, W., and Safran-Naveh, G., editors, *The Formal Complexity of Natural Language*. Springer, Dordrecht.

Peters, S. and Westerståhl, D. 2006. *Quantifiers in Language and Logic*. Oxford University Press.

Pickel, B. and Rabern, B. 2016. The antinomy of the variable: A tarskian resolution. *Journal of Philosophy*, 113(3):137–170.

Pickel, B. and Rabern, B. 2017. Does semantic relationism solve frege's puzzle? *Journal of Philosophical Logic*, 46:97–118.

Pietroski, P., Lidz, J., Hunter, T., and Halberda, J. 2009. The meaning of 'most' : Semantics, numerosity and psychology. *Mind & Language*, 24(5):554–585.

Pinillos, N. Á. 2011. Coreference and meaning. *Philosophical Studies*, 154(2):301–324.

Quine, W. 1951. *Mathematical Logic: Revised edition*. Harvard University Press.

Quine, W. 1960a. Variables explained away. In *Proceedings of the American Philosophical Society*, volume 104, pages 343–347. American Philosophical Society.

Quine, W. 1960b. *Word and Object*. Cambridge, MA, USA: MIT Press.

Quine, W. V. O. 1970. *Philosophy of Logic*. Harvard University Press.

Rabern, B. 2012. Against the identification of assertoric content with compositional value. *Synthese*, 189(1):75–96.

Rabern, B. 2013. Monsters in kaplan's logic of demonstratives. *Philosophical Studies*, 164:393–404.

Reinhart, T. 1997. Quantifier scope: How labor is divided between qr and choice functions. *Linguistics and philosophy*, pages 335–397.

Roberts, C. 1989. Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy*, 12:683–721.

Russell, B. 1903. *The principles of mathematics*. Norton Publishing.

Russell, B. 1905. On denoting. *Mind*, 14(56):479–493.

Salmon, N. 2006. A theory of bondage. *Philosophical Review*, 115(4):415–448.

Schlenker, P. 2006. Ontological symmetry in language: A brief manifesto. *Mind & language*, 21(4):504–539.

Schwarzschild, R. 2002. Singleton indefinites. *Journal of semantics*, 19(3):289–314.

Segerberg, K. 1973. Two-dimensional modal logic. *Journal of Philosophical Logic*, 2(1):77–96.

Shan, C. 2007. Lingustic side effects. In Barker, C. and Jacobson, P., editors, *Direct Compositionality*. Oxford University Press.

Stalnaker, R. 1975. Indicative conditionals. *Philosophia*, 5:269–286.

Stalnaker, R. 1978. Assertion. In Cole, P., editor, *Syntax and Semantics 9: Pragmatics*. New York: Academic Press.

Stalnaker, R. 1998. On the representation of context. In *Semantics and Linguistic Theory*, volume 6, pages 279–294.

Stanley, J. and Szabó, Z. G. 2000. On quantifier domain restriction. *Mind & Language*, 15:219–261.

Steedman, M. 2000. *The Syntactic Process*. MIT Press.

Stojnić, U., Stone, M., and Lepore, E. 2017. Discourse and logical form: Pronouns, attention and coherence. *Linguistics and Philosophy*, 40(5):519–547.

Strawson, P. F. 1950. On referring. *Mind*, 59(235):320–344.

Szabó, Z. G. 2000. Descriptions and uniqueness. *Philosophical Studies*, 101(1):29–57.

Szabó, Z. G. 2003. Definite descriptions without uniqueness: A reply to abbott. *Philosophical Studies*, pages 279–291.

Szabolcsi, A. 2010. *Quantification*. Reasearch Surveys in Linguistics. Cambridge University Press.

Szymanik, J. 2016. *Quantifiers and cognition: Logical and computational perspectives*, volume 96 of *Studies in Linguistics and Philosophy*. Springer.

Tarski, A. 1935. The concept of truth in formalized languages. In *Logic, Semantics, Metamathematics, papers for 1923 to 1938*. Hackett Publishing Company.

Troelstra, A. S. and Schwichtenberg, H. 2000. *Basic proof theory*. Number 43 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press.

Unger, C. 2012. Dynamic semantics as monadic computation. In *JSAI International Symposium on Artificial Intelligence*. Springer.

van den Berg, M. 1996. Dynamic generalized quantifiers. ms.

Van Eijck, J. and Kamp, H. 1997. Representing discourse in context. In *Handbook of logic and language*, pages 179–237. Elsevier.

Van Heijenoort, J. 1967. Logic as calculus and logic as language. *Synthese*, pages 324–330.

Vermeulen, C. 1995. Merging without mystery or: Variables in dynamic semantics. *Journal of Philosophical Logic*, 24:405–450.

Von Fintel, K.-U. 1994. *Restrictions on quantifier domains*. PhD thesis, University of Massachusetts Amherst.

Wadler, P. 1995. Monads for functional programming. In Jeuring, J. and Meijer, E., editors, *Advanced functional programming*, pages 24–52. Springer.

Wehmeier, K. 2018. The proper treatment of variables in predicate logic. *Linguistics and Philosophy*, 41:209–249.

Winter, Y. 1997. Choice functions and the scopal semantics of indefinites. *Linguistics and philosophy*, pages 399–467.

Yli-Vakkuri, J. 2013. Propositions and compositionality. *Philosophical Perspectives*.

Zimmerman, T. E. 1998. Remarks on the epistemic role of discourse referents. In *Logic, Language, and Computation*.

Zimmermann, T. E. and Sternefeld, W. 2013. *Introduction to Semantics: An Essential Guide to the Composition of Meaning*. De Gruyter Mouton.